Forecasting Emergency Department Crowding: An External, Multicenter Evaluation

Nathan R. Hoot, MD, PhD
Stephen K. Epstein, MD, MPP
Todd L. Allen, MD
Spencer S. Jones, PhD
Kevin M. Baumlin, MD
Neal Chawla, MD
Anna T. Lee
Jesse M. Pines, MD, MBA
Amandeep K. Klair, MD
Bradley D. Gordon, MD, MS
Thomas J. Flottemesch, PhD
Larry J. LeBlanc, PhD
Ian Jones, MD
Scott R. Levin, PhD
Chuan Zhou, PhD
Cynthia S. Gadd, PhD, MBA
Dominik Aronsky, MD, PhD

From the Vanderbilt University Medical Center, Nashville, TN (Hoot, I. Jones, Zhou, Gadd, Aronsky); Beth Israel Deaconess Medical Center, Boston, MA (Epstein); Intermountain Healthcare, Salt Lake City, UT (Allen); University of Utah, Salt Lake City, UT (S. S. Jones); Mount Sinai School of Medicine, New York, NY (Baumlin, Chawla, Lee); Hospital of the University of Pennsylvania, Philadelphia, PA (Pines, Klair); Regions Hospital, St. Paul, MN, and HealthPartners Research Foundation, Bloomington, MN (Gordon, Flottemesch); Owen Graduate School of Management, Nashville, TN (LeBlanc); and the Johns Hopkins University School of Medicine, Baltimore, MD (Levin).

Study objective: We apply a previously described tool to forecast emergency department (ED) crowding at multiple institutions and assess its generalizability for predicting the near-future waiting count, occupancy level, and boarding count.

Methods: The ForecastED tool was validated with historical data from 5 institutions external to the development site. A sliding-window design separated the data for parameter estimation and forecast validation. Observations were sampled at consecutive 10-minute intervals during 12 months (n=52,560) at 4 sites and 10 months (n=44,064) at the fifth. Three outcome measures—the waiting count, occupancy level, and boarding count—were forecast 2, 4, 6, and 8 hours beyond each observation, and forecasts were compared with observed data at corresponding times. The reliability and calibration were measured following previously described methods. After linear calibration, the forecasting accuracy was measured with the median absolute error.

Results: The tool was successfully used for 5 different sites. Its forecasts were more reliable, better calibrated, and more accurate at 2 hours than at 8 hours. The reliability and calibration of the tool were similar between the original development site and external sites; the boarding count was an exception, which was less reliable at 4 of 5 sites. Some variability in accuracy existed among institutions; when forecasting 4 hours into the future, the median absolute error of the waiting count ranged between 0.6 and 3.1 patients, the median absolute error of the occupancy level ranged between 9.0% and 14.5% of beds, and the median absolute error of the boarding count ranged between 0.9 and 2.8 patients.

Conclusion: The ForecastED tool generated potentially useful forecasts of input and throughput measures of ED crowding at 5 external sites, without modifying the underlying assumptions. Noting the limitation that this was not a real-time validation, ongoing research will focus on integrating the tool with ED information systems. [Ann Emerg Med. 2009;54:514-522.]

Provide feedback on this article at the journal's Web site, www.annemergmed.com.

0196-0644/$-see front matter

doi:10.1016/j.annemergmed.2009.06.006

Editor's Capsule Summary

What is already known on this topic

Emergency department (ED) census and workload can be highly variable. Accurate prediction of short-term future workload might improve efficiency if staffing could be tailored to the impending load.

What question this study addressed

Can readily available historical information be modeled to predict near-term ED operating conditions?

What this study adds to our knowledge

In this 5-site validation of the ForecastED model using historical data, the correlation between forecasted and actual operating conditions (waiting count, occupancy level, and boarding count) was reasonably good for a forecast of 2 hours in the future, but accuracy decreased as the prediction time increased beyond 2 hours into the future.

How this might change clinical practice

This study will not change clinical practice until it is demonstrated that this type of modeling can be used to improve the efficiency of care.

INTRODUCTION

Background

The emergency department (ED) serves essential needs in society, delivering emergency health care and simultaneously acting as a safety net provider.[1,2] The annual number of ED visits in the United States increased from 86.7 million in 1990 to 114.8 million in 2005.[3] During the same period, the number of EDs decreased from 5,172 to 4,611.[3] Moreover, 47% of US hospitals reported that they were operating at or over their ED capacity in 2007.[3] These divergent trends of capacity versus utilization may threaten the role of the ED both internationally and in the United States.[4-10] The Institute of Medicine reported that crowding has led the emergency medical system to reach "the breaking point."[11]

No universal consensus exists for the definition of "crowding" in the ED setting[12]; however, it may be described as a mismatch between patient demand for services and provider supply of resources. A portion of the recent literature has focused on techniques to monitor[13-17] and forecast[18-22] crowding, using various definitions. A recent white paper described management approaches that could be facilitated by such techniques in an effort to reduce crowding.[23] These include the 1-bed-ahead strategy, whereby inpatient units continuously anticipate the next patient requiring admission. More flexible staffing could also be implemented, whereby an ED schedules more nursing than necessary on average and allows shifts to end early during periods of low anticipated demand. Despite the possible applications, forecasting tools have not yet seen widespread operational adoption.

Importance

One challenge associated with monitoring and forecasting ED crowding is the generalization of models beyond the institutions where they were developed.[24-26] The decrease in predictive ability commonly seen when transporting models between sites may be due to the varying definitions of crowding, organizational structures, or workflow paradigms that exist among EDs. We recently described a discrete event simulation to forecast near-future ED crowding.[22] This tool, called ForecastED, was applied to predict crowding according to 7 input, throughput, and output measures[27] at the development site. We designed the model with generalizability as a central goal; however, its ability to forecast crowding at other sites has not been shown. Several steps are required to transform a prediction rule into a clinical decision rule that alters patient care; these include derivation, narrow validation, broad validation, and subsequently impact analysis.[28] The need for broad validation of the ForecastED tool motivated the present study.

Goal of This Investigation

The objective of this study was to externally validate the ForecastED tool for predicting near-future measures of crowding at institutions that are distinct from the original development site. More specifically, we intend to demonstrate whether the model parameters can be fitted for external sites without changing the underlying assumptions and determine whether the forecasting performance is comparable between external sites and the original development site.

MATERIALS AND METHODS

Theoretical Model of the Problem

The ForecastED tool implements a computerized "virtual ED" through a discrete event simulation intended to mimic the operations of an actual ED.[22] The process of developing the underlying model was theoretical and based on clinical experience. An interdisciplinary team proceeded iteratively within design constraints of forecasting power, minimal input requirements, and fast execution to determine a set of mathematical assumptions, together with software implementing them, specifying the operational structure of a generic ED.

The motivation to use discrete event simulation in forecasting, instead of another technique such as time series regression, is that it operates on patient data at a relatively detailed level, rather than at an institutional summary level; this granularity allows the tool to generate crowding forecasts without being limited to predetermined outcome measures. An autoregressive model would generate forecasts for a single outcome measure—for instance, the waiting count, the

Table 1. General characteristics of the participating validation sites.

Characteristics	Site A	Site B	Site C*	Site D	Site E
Medical center factors					
Inpatient capacity (# of beds)	702	425	520	1,171	620
Critical care capacity (# of beds)	92	50	70	133	77
Trauma service (accreditation)	Level I	Level I	Level I	N/A	Level I
Population served (adult/pediatric)	Adult	Both	Adult	Both	Adult
Metropolitan setting (urban/rural)	Urban	Urban	Urban	Urban	Urban
Emergency department factors					
Licensed capacity (# of beds)	25	35	25	34	39
Observation unit capacity (# of beds)	8	N/A	N/A	5	8
Triage system categories (# of levels)	4	5 (ESI)	5 (ESI)	5 (ESI)	5 (ESI)
Attending physicians (# on staff)	30	30	18	33	37
Operational data for 2006†					
Total volume (# of patients)	54,611	62,219	40,193	53,904	49,794
Diversion frequency (% of time)	7.0	1.0	0.3	3.6	2.4
Proportion eloped‡ (% of patients)	5.2	3.2	0.3	3.2	1.4
Proportion admitted (% of patients)	22.5	19.1	22.8	30.5	34.2
Waiting count (# of patients)	6 (2, 12)	3 (1, 7)	0 (0, 1)	3 (1, 6)	2 (1, 5)
Occupancy level (% of beds)	91 (73, 106)	77 (57, 91)	52 (32, 72)	85 (59, 113)	79 (62, 98)
Boarding count (# of patients)	8 (5, 11)	2 (1, 3)	1 (0, 2)	9 (5, 14)	12 (9, 16)

ESI, Emergency Severity Index.
*All operational data from site C were calculated according to 10 months of patient visits (March 1, 2006, to December 31, 2006), and the total volume was adjusted by a factor of 6/5 to extrapolate for the calendar year.
†Operational data are presented as counts, percentages, or median (interquartile range) as appropriate.
‡The proportion eloped is defined as the percentage of patients who register in the ED and leave before being assessed by a physician.

occupancy level, or the boarding count—that must be selected during model development. By comparison, the flexibility of discrete event simulation could allow one model to forecast all of these crowding indicators, among others. This property exists because the model input is a detailed list of patients who are in the ED at the present, and the model output is a detailed list of patients who are projected to be in the ED at a specified point in the near future.

Two design decisions may allow for generalizability of the ForecastED tool: First, specific numeric parameters for each statistical distribution are not built into it; these parameters are flexible and may change before the simulation is run. Second, institutional constants such as the number of ED beds or the number of acuity levels are likewise not built into it. In summary, the structural assumptions underlying the ForecastED tool were conserved among all validation sites, whereas the numeric parameters and institutional constants were intentionally allowed to vary between sites. The assumptions, parameters, and constants have been described in detail previously.[22]

Study Design

We validated the ForecastED tool using historical data from consecutive patient encounters during a 15-month period (November 1, 2005, to January 31, 2007) at 5 locations, all of which were geographically and operationally distinct from the location at which ForecastED was developed. The institutional review board at each participating site approved the study.

We maintained unique numeric parameters and institutional constants for each site, using a sliding-window study design, which was applied previously during the initial development and single-site validation of the ForecastED tool.[22] Given an observation time and institution, this technique uses data from the recent past—4 weeks in the present study—to fit parameters for all statistical distributions within the simulation. Then it uses data from the near future—2, 4, 6, and 8 hours in the present study—for forecast validation. These windows are always relative and are adjusted each time the observation point is moved. The primary purpose of this design was to ensure that the sets of data used to estimate the parameters and to validate the forecasts remained independent at all times. The secondary purpose of this design was to keep the parameters accurate with respect to seasonal variation that may occur at a given institution.

We generated observations in time series at 10-minute intervals between January and December 2006 (n=52,560) at sites A, B, D, and E. At site C, we repeated this process at 10-minute intervals between March and December 2006 (n=44,064) because of the unavailability of data from earlier dates.

Setting

Our study took place at 5 academic, urban, tertiary care medical centers. Three of the sites are located in the northeastern region of the United States, whereas the other 2 are located in the midwestern and western regions. General descriptive characteristics for each participating institution and its affiliated ED are presented in Table 1. The hospitals range in size from 425 to 1,171 beds, including 50 to 133 critical care beds. Four of the sites are designated as Level I trauma centers.

Three serve adult populations, whereas 2 serve both adult and pediatric populations. The EDs range in licensed capacity from 25 to 39 beds. Three of the institutions have observation units managed by the ED that range in size from 5 to 8 beds. Because the EDs at these sites can access the observation unit for overflow when crowded, these beds are included in the total ED capacity for the purpose of the simulation. Thus, the total ED capacity specified for the ForecastED tool was 33 beds at site A, 35 beds at site B, 25 beds at site C, 39 beds at site D, and 47 beds at site E. Four of the sites triage ED patients with the Emergency Severity Index, a 5-level score in which lower values indicate greater urgency[29]; one site triages ED patients according to a 4-level ranking. The number of ED attending physicians employed by the participating sites ranges from 18 to 37. Local policy at each of the 5 sites allows for ambulance diversion during periods of crowding.

Selection of Participants

The study included data from all patient visits at each participating site during the study period, with 3 exclusion criteria applied: (1) Patient visits were excluded if the time of registration or the time of discharge was missing because we could not accurately determine when the patient was present in the ED. (2) Patient visits were excluded if the patient was admitted directly to the hospital without being placed into an ED bed because these patients tend not to compete for ED resources. Such patient encounters are referred to as "immediate admissions" in this study. (3) At site A, which has a separate psychiatric ED, patient visits were excluded if the chief complaint was purely psychiatric because these patients were not deemed to contribute much to crowding at that site.

Data Collection and Processing

The following variables are required for each patient visit to estimate parameters for, and to generate forecasts by, the ForecastED tool[22]: (1) time of initial registration at the ED, (2) time placed into an ED treatment bed, (3) time of hospital bed request, if applicable, (4) time of discharge from the ED facility, (5) triage category assigned to the patient, (6) whether the patient left without being seen, and (7) whether the patient was admitted to the hospital. Each institution collected these data from ED patient-tracking information systems. Two sites used commercial information systems, whereas 3 sites used information systems that were developed in house. The following institutional constants were also supplied to the model as necessary to generate forecasts[22]: total ED capacity, including licensed treatment beds and, where applicable, beds within an ED-managed observation unit; and number of acuity levels in the ranking system used to triage ED patients.

Forecasts were obtained at a given time and institution by using the following series of steps: (1) All patients who were discharged from the ED during the preceding 4 weeks were identified. (2) The required parameters for each statistical distribution governing the simulation were estimated with the set of historical patient encounters. Formulas used for parameter estimation are given in Appendix E1 (available online at http://www.annemergmed.com). (3) All patients who were present in the ED at the observation time of interest were identified. (4) The simulation was initialized according to the set of current patient encounters, with patients being placed in the virtual waiting room or virtual ED beds as appropriate. (5) The simulation ran 2 hours into the future before terminating. At that time, the state of the virtual ED was noted, and the waiting count, occupancy level, and boarding count were measured. (6) The previous 2 steps were repeated 1,000 times to obtain an average for each outcome measure. (7) The previous 3 steps were repeated with the simulation running 4, 6, and 8 hours into the future. The actions of data processing and parameter estimation were automated with a Python language script (version 2.3.5; available online at http://www.python.org).

Outcome Measures

The waiting count was defined as the number of patients in the waiting room. The occupancy level was defined as the total number of patients in ED beds divided by the number of treatment beds (this value may exceed 100% when patients are treated in nonlicensed areas such as hallway beds or chairs). The boarding count was defined as the number of patients with hospital admission orders who await inpatient beds. These 3 outcome measures, with identical definitions, were used during the development of ForecastED.[22] Details on calculating these data with raw patient information are provided in Appendix E1 (available online at http://www.annemergmed.com).

Primary Data Analysis

We calculated the Pearson's *r* coefficient of correlation to quantify the reliability of the forecasts with respect to the actual operational measure at the corresponding point in the future. For example, when the simulation used the operational status at noon to forecast the occupancy level 8 hours in the future, we compared the resulting forecast against the known actual occupancy level at 8 PM. The Pearson's *r* value reflects the degree of linear association between 2 measures, without penalizing for any consistent numeric bias. The square of this value describes the proportion of total variation explained by the forecasts. We calculated the Pearson's *r* with 95% confidence intervals (CIs) using 250 iterations of the ordinary nonparametric bootstrap method.[30]

We recognized that the crowding status of an ED is likely to be autocorrelated. For example, the occupancy level at noon on a given day provides some potentially useful information about the occupancy level at 1 PM. We considered the present status of an ED to be a naive predictor of the near-future status of that same ED, so we used this as our control measure for describing the additional utility provided by the forecasts. The autocorrelation gives the Pearson's *r* coefficient of correlation within a single time series, such that one point in time is compared with a later point in time, following a specified interval. The usefulness of the simulation forecasts was judged by whether the reliability of the simulation forecasts exceeded the inherent autocorrelation within each actual operational measure. We calculated the autocorrelation coefficients with

95% CIs by using 250 iterations of the ordinary nonparametric bootstrap method.[30]

Although correlation coefficients are useful to assess the amount of collinearity that exists between 2 measures, they cannot detect any systematic bias that may exist; that is, any consistent overestimation or underestimation that would reduce numeric agreement.[31] To assess whether a systematic bias existed between the simulation forecasts and the actual operational measures, we calculated the mean and SD of the residual error. We would consider a mean near zero, in the context of the associated SD, to demonstrate good calibration.

The above statistical analysis was identical to the protocol used during the initial, single-center validation of ForecastED.[22] We performed one additional step of measuring the accuracy, with the goal of making the results more easily interpretable. First, we calibrated the forecasts with the best-fitting line between predicted and actual operational data. This line was calculated with 7 days of time-series data (n=1,008) preceding each observation, and the resulting linear transformation was used to obtain a single bias-corrected forecast. The calibration process was repeated over each time series of forecasts, in a manner analogous to the sliding-window study design described above. This step was justified on the grounds that a systematic bias was found to exist during previous research on the ForecastED tool,[22] and it mirrors the intended real-world application of the tool. Next, we calculated the median absolute error between the calibrated forecasts and the actual operational data, with values closer to zero denoting greater accuracy.

We conducted all statistical analyses with R (version 2.8.1; available online at http://www.r-project.org).

RESULTS

Summary statistics of the conditions within each participating ED during the study period are presented in Table 1. The ED at site C was the least crowded in terms of the total volume (40,193 visits), percentage of patients leaving without being seen (0.3%), median number of patients in the waiting room (0 patients), median percentage of beds occupied (52%), and median number of patients boarding simultaneously (1 patient). The ED at site B had the highest annual volume (62,219 visits), whereas the ED at site A had the highest percentage of patients leaving without being seen (5.2%), median number of patients in the waiting room (6 patients), and median percentage of beds occupied (91%). The ED at site E boarded the highest median number of patients simultaneously (12 patients). The percentage of total time spent on ambulance diversion ranged among the participating sites from 0.3% at site C to 7.0% at site A. For comparison, the US average of total time spent on diversion in 2002 was 2.9%, or 7.6% among centers with high annual volumes.[32]

The number of patient visits excluded from the analysis totaled 1,418 patient visits from site A (0.2% immediate admissions, 1.9% purely psychiatric), 1,035 patient visits from site B (1.3% missing data, <0.1% immediate admissions), 28 patient visits from site C (<0.1% immediate admissions), 0 patient visits from site D, and 14 patient visits from site E (<0.1% immediate admissions). After calculation of the time series of simulation forecasts and operational measures, implausible values were found to exist that were associated with episodes of computer system downtime at sites B and C. We removed affected observations from the time series before conducting statistical analysis; these included 2 segments of 36 hours each at site B and 4 segments of 24 hours each at site C.

The numeric agreement between the observed occupancy level and the simulation forecasts at site A may be visualized with the scatterplots in the Figure. The full set of data representing all outcome measures and participating institutions may be visualized in Appendix E2 (available online at http://www.annemergmed.com).

The site-by-site reliability, calibration, and accuracy of the forecasts 4 hours in the future are presented in Tables 2, 3, and 4, respectively. The full results on the forecasting performance at 2, 4, 6, and 8 hours in the future are presented in Appendix E3 (available online at http://www.annemergmed.com). Three trends were observed in the results for all operational measures considered. First, the forecasts became less reliable as the forecasts extended farther into the future. For example, the forecasts of the waiting room count at site A had Pearson's r statistics of 0.79, 0.71, 0.65, and 0.61 with the observed waiting room count 2, 4, 6, and 8 hours in the future, respectively. Second, the reliability of the simulation tended to decrease more rapidly as the forecasting length increased, compared with the inherent autocorrelation which decreased more rapidly. For example, the forecasts of the occupancy level at site D had Pearson's r statistics of 0.87, 0.79, 0.76, and 0.73 with the observed occupancy level 2, 4, 6, and 8 hours in the future, respectively; the inherent autocorrelation was 0.86, 0.61, 0.31, and 0.03 when calculated for delays of 2, 4, 6, and 8 hours, respectively. Third, there was sufficient systematic bias, as indicated by the residual mean differing from zero, to justify an additional calibration step in real-world deployment. For example, when forecasting the boarding count at site B, the model tended to overestimate by 1.5±1.7, 2.0±1.9, 2.4±2.0, and 2.7±2.1, respectively, at 2, 4, 6, and 8 hours in the future.

The forecasts of the waiting count were more reliable than the inherent autocorrelation at 4 of 5 sites. At sites B, D, and E, the median absolute error was consistently accurate within 1.5 to 2.1 patients at 2 hours in the future and within 1.7 to 2.5 patients at 8 hours in the future. At site A, the median absolute error was slightly less accurate, within 2.7 patients at 2 hours in the future and 3.5 patients at 8 hours in the future. This may be explained by the observation that site A had a busier waiting room, with a median of 6 patients and 75th percentile of 12 patients in the waiting room. At site C, the forecasts of the waiting count correlated weakly with the observed waiting count across all forecasting lengths (r=0.15 at 2, 4, 6, and 8 hours). However, the accuracy at site C remained constant at 0.6 patients up to 8 hours in the future.

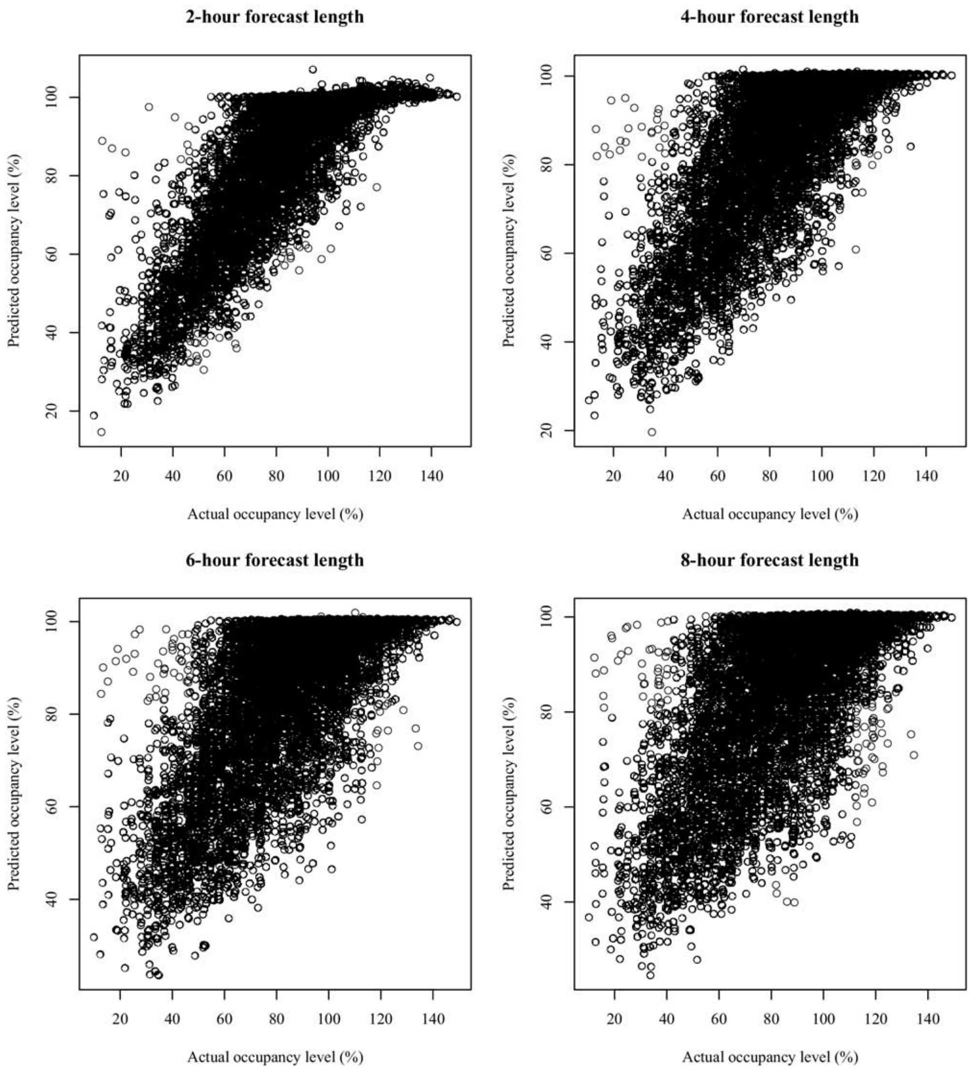

Figure. Actual (*x* axis) versus predicted (*y* axis) occupancy levels at site A. The data represent forecast lengths of 2 hours (top left), 4 hours, (top right), 6 hours (bottom left), and 8 hours (bottom right). The occupancy level is expressed as the percentage of licensed beds that are filled at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

Table 2. Reliability of the simulation versus autocorrelation in forecasting operational data, 4 hours in the future.*

	Site A	Site B	Site C	Site D	Site E
Waiting count					
Simulation (*r*)	0.71 (0.70-0.71)	0.52 (0.51-0.53)	0.15 (0.14-0.16)	0.56 (0.56-0.57)	0.65 (0.64-0.66)
Autocorrelation (*r*)	0.50 (0.49-0.51)	0.45 (0.44-0.46)	0.08 (0.06-0.09)	0.41 (0.40-0.42)	0.44 (0.44-0.45)
Occupancy level					
Simulation (*r*)	0.78 (0.77-0.78)	0.80 (0.80-0.81)	0.77 (0.77-0.77)	0.79 (0.79-0.79)	0.83 (0.83-0.83)
Autocorrelation (*r*)	0.52 (0.51-0.52)	0.46 (0.45-0.47)	0.37 (0.36-0.37)	0.61 (0.61-0.62)	0.60 (0.60-0.61)
Boarding count					
Simulation (*r*)	0.72 (0.72-0.73)	0.41 (0.40-0.41)	0.26 (0.25-0.27)	0.70 (0.70-0.71)	0.57 (0.57-0.58)
Autocorrelation (*r*)	0.73 (0.73-0.74)	0.48 (0.47-0.49)	0.17 (0.16-0.18)	0.65 (0.64-0.65)	0.68 (0.68-0.68)

*The Pearson coefficient of correlation is presented, with lower and upper bounds of the 95% CI in parentheses.

The occupancy level forecasts exceeded the inherent autocorrelation at all participating sites. At sites A, B, C, and E, the median absolute error of the occupancy level forecasts ranged from 7.1% to 9.6% of beds at 2 hours in the future, to 10.0% to 12.0% of beds at 8 hours in the future. At site D, the forecasts were less accurate, increasing from 11.6% of beds to

Table 3. Calibration of the simulation in forecasting operational data, 4 hours in the future.*

	Site A	Site B	Site C	Site D	Site E
Waiting count (# of patients)	0.6 ± 6.9	−2.5 ± 4.1	−0.3 ± 1.2	3.6 ± 9.2	1.2 ± 5.2
Occupancy level (% of beds)	−0.3 ± 15.4	−0.4 ± 13.4	7.0 ± 16.9	−5.6 ± 23.5	2.6 ± 13.3
Boarding count (# of patients)	0.2 ± 3.0	2.0 ± 1.9	1.0 ± 1.5	−1.2 ± 4.4	−2.0 ± 4.4

*The forecasting residuals are summarized with the mean±SD.

Table 4. Absolute error of the simulation in forecasting operational data, 4 hours in the future.*

	Site A	Site B	Site C	Site D	Site E
Waiting count (# of patients)	3.1 (1.6-5.1)	2.4 (1.3-3.7)	0.6 (0.4-0.7)	2.0 (1.0-3.2)	1.6 (0.8-2.7)
Occupancy level (% of beds)	10.4 (4.9-17.5)	9.0 (4.3-15.4)	11.1 (5.2-19.1)	14.5 (6.9-25.0)	9.0 (4.3-15.4)
Boarding count (# of patients)	1.9 (0.9-3.2)	1.0 (0.5-1.6)	0.9 (0.4-1.2)	2.8 (1.3-4.7)	2.7 (1.3-4.5)

*The median absolute error is presented, with the interquartile range in parentheses.

16.8% of beds, respectively, at 2 hours and 8 hours in the future. The scatterplot of the actual and predicted occupancy levels at site D, shown in Figure E11 of Appendix E2 (available online at http://www.annemergmed.com), shows an unusual trend: the forecasts of the occupancy level tend to be capped near 100%, yet site D tends to reach occupancy levels exceeding 150% more often than the other sites.

Considering the forecasts of the boarding count, the reliability at site D exceeded the inherent autocorrelation, but this did not hold true for any other site. At sites B and C, which boarded the fewest patients at one time among the 5 participating institutions, the median absolute error of the boarding count forecasts remained close to 1.0 patient up to 8 hours in the future. At sites A, D, and E, the median absolute error ranged from 1.4 to 2.0 patients at 2 hours in the future to 2.3 to 3.8 patients at 8 hours in the future.

LIMITATIONS

There are a number of limitations to our study, including the use of historical patient data. Although the ForecastED tool is intended for real-time application, we considered it technically more feasible to validate the tool at 5 institutions with offline analysis instead of live deployment. The differences between nonconcurrent and concurrent data cleaning may have affected the results. The technique of cleaning data in real time is more difficult because of inherent uncertainty of the information, so an accurate determination cannot always be made about whether a given patient meets the inclusion criteria. For example, when data are examined from a center at which psychiatric patients should be excluded, all patients having solely psychiatric complaints can be identified in retrospect. However, when patients are considered in real time as they present to triage, one cannot immediately know who will be referred to the psychiatric ED. Because of this, the results may be considered optimistic for each participating site; however, the percentages of excluded data were small, so it seems unlikely that the real-time performance would degrade substantially because of this effect. The results may still provide useful information about the generalizability of the ForecastED tool.

A selection bias may exist in the study, with respect to which institutions participated; the 5 sites were chosen with an intentional focus on academic, metropolitan EDs. Thus, the validation results may not be representative of rural or nonteaching hospitals. We justify the sampling of participating sites on the grounds that the crowding burden disproportionately affects EDs such as the ones described here,[3,33] and a method to forecast crowding offers more potential value to busy, crowded EDs. An additional, practical constraint required all participating institutions to have computerized patient tracking systems. This research would have been difficult to perform otherwise, given that several data elements needed to be obtained for each patient during a lengthy period.

Many simplifying assumptions were made in the process of applying the ForecastED tool across 5 institutions. The discrete event simulation represents a generic ED, and it does not capture the rich variety of strategies that physicians and administrators may use to improve ED operations. It treats all beds identically, so it does not handle trauma bays differently from unmonitored beds. It does not model fast-track beds, which may be open only during specified hours of the day and are not equipped to handle patients with severe conditions. The schemes used to allocate beds in the ED and hospital are simple; this may explain why forecasts of the occupancy level rarely exceed 100% and why forecasts of the boarding count do not exceed the autocorrelation in reliability. In its present form, the simulation does not capture the different roles that physicians may play during a shift. The ForecastED tool was designed to represent the common denominator across institutions, and despite this, it performed generally well in forecasting the waiting count and the occupancy level.

DISCUSSION

We successfully estimated parameters for, and obtained forecasts from, the ForecastED tool at 5 institutions. All required numeric parameters were fitted with historical data from ED information systems, and no modification to the

model's assumptions, as originally described, was necessary to generate forecasts in different settings.[22]

The results show that, with respect to its ability to forecast the waiting count and the occupancy level, the ForecastED tool generalizes fairly well among the participating institutions. The degrees of reliability and accuracy varied among sites; however, after controlling for the intrinsic difficulty of forecasting operational data as measured by the autocorrelation, the simulation gave additional predictive information. Similar observations were made at the original development site of the ForecastED tool.[22] The forecasts of the occupancy level appeared to remain useful up to 8 hours in the future, whereas the forecasts of the waiting count may be most useful 4 hours or fewer in the future. The pattern observed in the occupancy level forecasts at site D suggests a future refinement to the ForecastED tool: The simulation assumes that an occupancy level of 100% is exceeded only for the most critically ill patients,[22] even though some institutions—notably, site D in this study—routinely place patients in hallways or double the occupancy in rooms, achieving occupancy levels of 150% or even 200%. This assumption may need to be relaxed in future research.

The ForecastED tool generally provided little additional predictive information, beyond the autocorrelation, for the number of boarding patients. This observation held true across 4 of the 5 external validation sites. This was not the case for the original development site,[22] and this difference may provide useful information about potential improvements for the ForecastED tool. It has been suggested that extended boarding of patients in the ED substantially exacerbates the crowding problem.[7,34-36] Furthermore, some researchers have reported that improving access to hospital beds may lessen the burden of ED crowding.[37-38] According to these 2 lines of reasoning, one might improve the forecasts of the boarding count by making the process of simulated hospital bed allocation more robust within the ForecastED tool. This might be achieved through the use of granular data from the affiliated hospital, such as the present inpatient occupancy and projected lengths of stay.

Numerous challenges are associated with forecasting ED crowding; for instance, the lack of uniformly accepted definitions for crowding and common operational data.[39,40] The ForecastED tool was developed with the goal of addressing the above issues, through generating forecasts in a target-agnostic manner.[22] The variability observed between sites in our study may be attributable to differences in operating room schedules, differences in policy for managing patient flow, and unique factors that influence input from the community. Substantial complexity exists among the hospital factors that affect crowding, and this may augment the complexity of forecasting tools in later research. The practical value of forecasting ED crowding depends on real-time data, which is not yet available in many centers; furthermore, this work could gain additional utility if regional information sharing within hospital networks became common practice.[41] Most importantly, any forecasts of ED crowding must be combined with a plan to intervene and alleviate crowding to achieve practical value; until this occurs, we refrain from speculating on whether the forecasts are sufficiently accurate.

In summary, we validated the ForecastED tool, without modifying any assumptions from the original description, to forecast ED crowding at 5 institutions separate from the development site. The tool generated potentially useful forecasts of input and throughput measures of ED crowding, and further opportunities may exist to improve the forecasts of output measures.[27] Future research will address the question of how to operationalize the ForecastED tool, with the goal of determining whether it can improve management of ED workflow to lessen the crowding burden.

Supervising editor: David J. Magid, MD, MPH

Author contributions: NRH and DA conceived the study. All authors contributed substantially to the study design. DA and IJ obtained research funding. SKE, TLA, SSJ, KMB, NC, ATL, JMP, AKK, BDG, and TJF collected the data. NRH and CZ performed the statistical analysis. NRH drafted the article, and all authors contributed substantially to its revision. NRH takes responsibility for the paper as a whole.

Funding and support: By *Annals* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article that might create any potential conflict of interest. See the Manuscript Submission Agreement in this issue for examples of specific conflicts covered by this statement. The first author was supported by National Library of Medicine grant LM07450 and National Institute of General Medical Studies grant T32 GM07347. The research was also supported by the National Library of Medicine grant R21 LM009002. This project is an academic endeavor. It was supported by federal funding, as noted above, and there was no corporate funding. There are no current plans to commercialize this research that would cause any conflicts of interest.

Publication dates: Received for publication June 4, 2008. Revisions received November 6, 2008; March 19, 2009; and May 20, 2009. Accepted for publication June 1, 2009. Available online August 28, 2009.

Reprints not available from the authors.

Address for correspondence: Nathan R. Hoot, MD, PhD, 400 Eskind Biomedical Library, 2209 Garland Ave, Nashville, TN 37232; 615-498-1979, fax 615-936-1427; E-mail nathan.hoot@vanderbilt.edu.

REFERENCES

1. Gordon JA, Billings J, Asplin BR, et al. Safety net research in emergency medicine: proceedings of the Academic Emergency Medicine Consensus Conference on "The Unraveling Safety Net." *Acad Emerg Med.* 2001;8:1024-1029.
2. American Academy of Pediatrics Committee on Pediatric Emergency Medicine. Overcrowding crisis in our nation's emergency departments: is our safety net unraveling? *Pediatrics.* 2004;114:878-888.

3. American Hospital Association. TrendWatch Chartbook 2007. American Hospital Association Web site. Available at: http://www.aha.org/aha/trendwatch/2007/cb2007chapter3.ppt. Accessed July 18, 2007.
4. Li G, Lau JT, McCarthy ML, et al. Emergency department utilization in the United States and Ontario, Canada. *Acad Emerg Med.* 2007;14:582-584.
5. Proudlove NC, Gordon K, Boaden R. Can good bed management solve the overcrowding in accident and emergency departments? *Emerg Med J.* 2003;20:149-155.
6. Espinosa G, Miró O, Sánchez M, et al. Effects of external and internal factors on emergency department overcrowding. *Ann Emerg Med.* 2002;39:693-695.
7. Fatovich DM, Nagree Y, Sprivulis P. Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia. *Emerg Med J.* 2005;22:351-354.
8. Shih FY, Ma MH, Chen SC, et al. ED overcrowding in Taiwan: facts and strategies. *Am J Emerg Med.* 1999;17:198-202.
9. Rehmani R. Emergency section and overcrowding in a university hospital of Karachi, Pakistan. *J Pak Med Assoc.* 2004;54:233-237.
10. Graff L. Overcrowding in the ED: an international symptom of health care system failure. *Am J Emerg Med.* 1999;17:208-209.
11. Committee on the Future of Emergency Care in the United States Health System. *Hospital-Based Emergency Care: At the Breaking Point.* Washington, DC: National Academies Press; 2006.
12. Hwang U, Concato J. Care in the emergency department: how crowded is overcrowded? *Acad Emerg Med.* 2004;11:1097-1101.
13. Bernstein SL, Verghese V, Leung W, et al. Development and validation of a new index to measure emergency department crowding. *Acad Emerg Med.* 2003;10:938-942.
14. Reeder TJ, Burleson DL, Garrison HG. The overcrowded emergency department: a comparison of staff perceptions. *Acad Emerg Med.* 2003;10:1059-1064.
15. Weiss SJ, Derlet R, Arndahl J, et al. Estimating the degree of emergency department overcrowding in academic medical centers: results of the National ED Overcrowding Study (NEDOCS). *Acad Emerg Med.* 2004;11:38-50.
16. Asplin BR, Rhodes KV, Flottemesch TJ, et al. Is this emergency department crowded? a multicenter derivation and evaluation of an emergency department crowding scale (EDCS) [abstract]. *Acad Emerg Med.* 2004;11:484-485.
17. Epstein SK, Tian L. Development of an emergency department work score to predict ambulance diversion. *Acad Emerg Med.* 2006;13:421-426.
18. Tandberg D, Qualls C. Time series forecasts of emergency department patient volume, length of stay, and acuity. *Ann Emerg Med.* 1994;23:299-306.
19. Asplin BR, Flottemesch TJ, Gordon BD. Developing models for patient flow and daily surge capacity research. *Acad Emerg Med.* 2006;13:1109-1113.
20. Jones SS, Thomas A, Evans RS, et al. Forecasting daily patient volumes in the emergency department. *Acad Emerg Med.* 2008;15:159-170.
21. McCarthy ML, Zeger SL, Ding R, et al. The challenge of predicting demand for emergency department services. *Acad Emerg Med.* 2008;15:337-346.
22. Hoot NR, LeBlanc LJ, Jones I, et al. Forecasting emergency department crowding: a discrete event simulation. *Ann Emerg Med.* 2008;52:116-125.
23. Eitel DR, Rudkin SE, Malvehy MA, et al. Improving service quality by understanding emergency department flow: a white paper and position statement prepared for the American Academy of Emergency Medicine. *J Emerg Med.* 2008 May 29. [Epub ahead of print.]
24. Jones SS, Allen TL, Flottemesch TJ, et al. An independent evaluation of four quantitative emergency department crowding scales. *Acad Emerg Med.* 2006;13:1204-1211.
25. Hoot NR, Zhou C, Jones I, et al. Measuring and forecasting emergency department crowding in real time. *Ann Emerg Med.* 2007;49:747-755.
26. McCarthy ML, Aronsky D, Jones ID, et al. The emergency department occupancy rate: a simple measure of emergency department crowding? *Ann Emerg Med.* 2008;51:15-24.
27. Asplin BR, Magid DJ, Rhodes KV, et al. A conceptual model of emergency department crowding. *Ann Emerg Med.* 2003;42:173-180.
28. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med.* 2006;144:201-209.
29. Wuerz RC, Milne LW, Eitel DR, et al. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med.* 2000;7:236-242.
30. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci.* 1986;1:54-77.
31. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1:307-310.
32. Burt CW, McCaig LF, Valverde RH. Analysis of ambulance transports and diversions among US emergency departments. *Ann Emerg Med.* 2006;47:317-326.
33. Andrulis DP, Kellermann A, Hintz EA, et al. Emergency departments and crowding in United States teaching hospitals. *Ann Emerg Med.* 1991;20:980-986.
34. Fromm RE Jr, Gibbs LR, McCallum WG, et al. Critical care in the emergency department: a time-based study. *Crit Care Med.* 1993;21:970-976.
35. Schull MJ, Lazier K, Vermeulen M, et al. Emergency department contributors to ambulance diversion: a quantitative analysis. *Ann Emerg Med.* 2003;41:467-476.
36. Schneider SM, Gallery ME, Schafermeyer R, et al. Emergency department crowding: a point in time. *Ann Emerg Med.* 2003;42:167-172.
37. Dunn R. Reduced access block causes shorter emergency department waiting times: an historical control observational study. *Emerg Med (Fremantle).* 2003;15:232-238.
38. McConnell KJ, Richards CF, Daya M, et al. Effect of increased ICU capacity on emergency department length of stay and ambulance diversion. *Ann Emerg Med.* 2005;45:471-478.
39. Solberg LI, Asplin BR, Weinick RM, et al. Emergency department crowding: consensus development of potential measures. *Ann Emerg Med.* 2003;42:824-834.
40. Asplin BR. Measuring crowding: time for a paradigm shift. *Acad Emerg Med.* 2006;13:459-461.
41. Sprivulis P, Gerrard B. Internet-accessible emergency department workload information reduces ambulance diversion. *Prehosp Emerg Care.* 2005;9:285-291.

APPENDIX E1. Estimating parameters in ForecastED.

Six random-number distributions govern the flow of patients through the ED, as conceptualized in the discrete event simulation. The following provides details on the standard maximum likelihood estimation formulas applied to update the parameters at each institution. All parameters were re-estimated to be specific for a given time and institution; however, the underlying mathematical assumptions were preserved between all times and institutions.

A nonstationary Poisson distribution described the process of patient arrivals to the ED. A total of 168 parameters, named in this paragraph as μ_1, μ_2, . . .μ_{168}, defined the distribution. Each parameter may be interpreted as the number of patient arrivals per hour, within each hour interval of the week. The parameter μ_1 described Sunday from midnight to 1 AM, the parameter μ_2 described Sunday from 1 AM to 2 AM, etc. To estimate the parameter μ_x, we used the formula $n_x/4$; in this formula, n_x denoted the number of patients having registration times that fell within the hour x during each of the preceding 4 weeks.

A logistic regression governed the process of patients leaving without being seen. A total of 2 parameters, named in this paragraph as α and β, defined the distribution. The α parameter was the constant of a 1-variable logistic regression, whereas the β parameter was the coefficient. The independent variable x in the equation was the number of patients in the waiting room at the time a given patient registered. The dependent variable y in the equation was a zero or 1, indicating whether a given patient left without being seen. For an unknown case, the dependent variable y provides the log-odds of a patient leaving without being seen, which, after application of the formula $e^y/(1+e^y)$, provides the corresponding probability. We used the standard method for fitting a logistic regression to estimate the values of α and β.

A multinomial distribution described the process of assigning patient acuity levels in triage. A total of n parameters, named in this paragraph as p_1, p_2, . . .p_n, defined the distribution, where n represents the number of acuity levels for an institution; at most sites, the value of n was 5, if the Emergency Severity Index (ESI) was used. Each parameter may be interpreted as the probability of a given patient belonging to acuity level x. To estimate the parameter p_x, we used the formula n_x/m; in this formula, n_x denoted the number of patients having acuity level x, and m denoted the total number of patients having a valid acuity level recorded.

A log normal distribution conditioned on the patient acuity level described the time of evaluation and treatment in the ED. A total of 2n parameters, named in this paragraph as μ_1, μ_2, . . .μ_n and σ_1, σ_2, . . .σ_n, defined the distribution, where n represents the number of acuity levels for an institution; at most sites, the value of n was 5, if the ESI was used. To estimate the parameter μ_x, we used the formula $\Sigma(\log t_{xi})/n_x$; in this formula, t_{xi} denoted the time of evaluation and treatment (as calculated by subtracting the discharge time from the bed placement time) for patient i having acuity level x, and n_x denoted the number of patients having acuity level x. To estimate the parameter σ_x, we used the square root of the formula $\Sigma[(\log t_{xi})-\mu_x]^2/n_x$; in this formula, the definitions of t_{xi}, μ_x, and n_x retain the previous definitions. These calculations excluded patients who were admitted to the hospital, to avoid confounding the evaluation and treatment times with boarding times.

A Bernoulli distribution conditioned on the patient acuity level described the process of hospital admission decisions within the ED. A total of n parameters, named in this paragraph as p_1, p_2, . . .p_n, defined the distribution, where n represents the number of acuity levels for an institution; at most sites, the value of n was 5, if the ESI was used. Each parameter may be interpreted as the probability of a given patient with acuity level x being admitted to the hospital. To estimate the parameter p_x, we used the formula n_x/m_x; in this formula, n_x denoted the number of patients having acuity level x who were admitted to the hospital (as identified by the disposition flag), and m_x denoted the total number of patients having acuity level x.

A nonstationary Poisson distribution described the process of hospital bed openings for the ED. A total of 168 parameters, named in this paragraph as μ_1, μ_2, . . .μ_{168}, defined the distribution. Each parameter may be interpreted as the number of boarding patients discharged from the ED into the hospital per hour, within each hour interval of the week. The parameter μ_1 described Sunday from midnight to 1 AM, the parameter μ_2 described Sunday from 1 AM to 2 AM, etc. To estimate the parameter μ_x, we used the formula $n_x/4$; in this formula, n_x denoted the number of boarding patients (as distinguished from nonboarding patients by the disposition flag) having discharge times that fell within the hour x during each of the preceding 4 weeks.

Calculating Outcomes From Patient-Level Data

As noted in the article, the raw output of the ForecastED simulation is a list of patients who may be present in the ED at a specified point t in the future. For each patient i, several values are specified, including the following: (1) time of initial registration at the ED, named as r_i; (2) time placed into an ED treatment bed, named as b_i; (3) time of hospital admission order, if applicable, named as a_i; and (4) time of discharge from the ED facility, named as d_i. Additional transformations are necessary to translate this information into operational data of interest. Patient i was identified as in the waiting room at time t if $r_i \leq t < b_i$, or if $r_i \leq t < d_i$ and patient i was never placed into an ED treatment bed. Patient i was identified as occupying an ED treatment bed at time t if $b_i \leq t < d_i$. Patient i was identified as awaiting hospital admission at time t if $a_i \leq t < d_i$. Using this information, the waiting count was calculated by counting the patients who were in the waiting room at time t. The occupancy level was calculated by counting the patient who were in ED treatment beds at time t, and then dividing by the total number of ED treatment beds. The boarding count was calculated by counting the patients who were awaiting hospital admission at time t.

APPENDIX E2. Site-by-site data visualization.

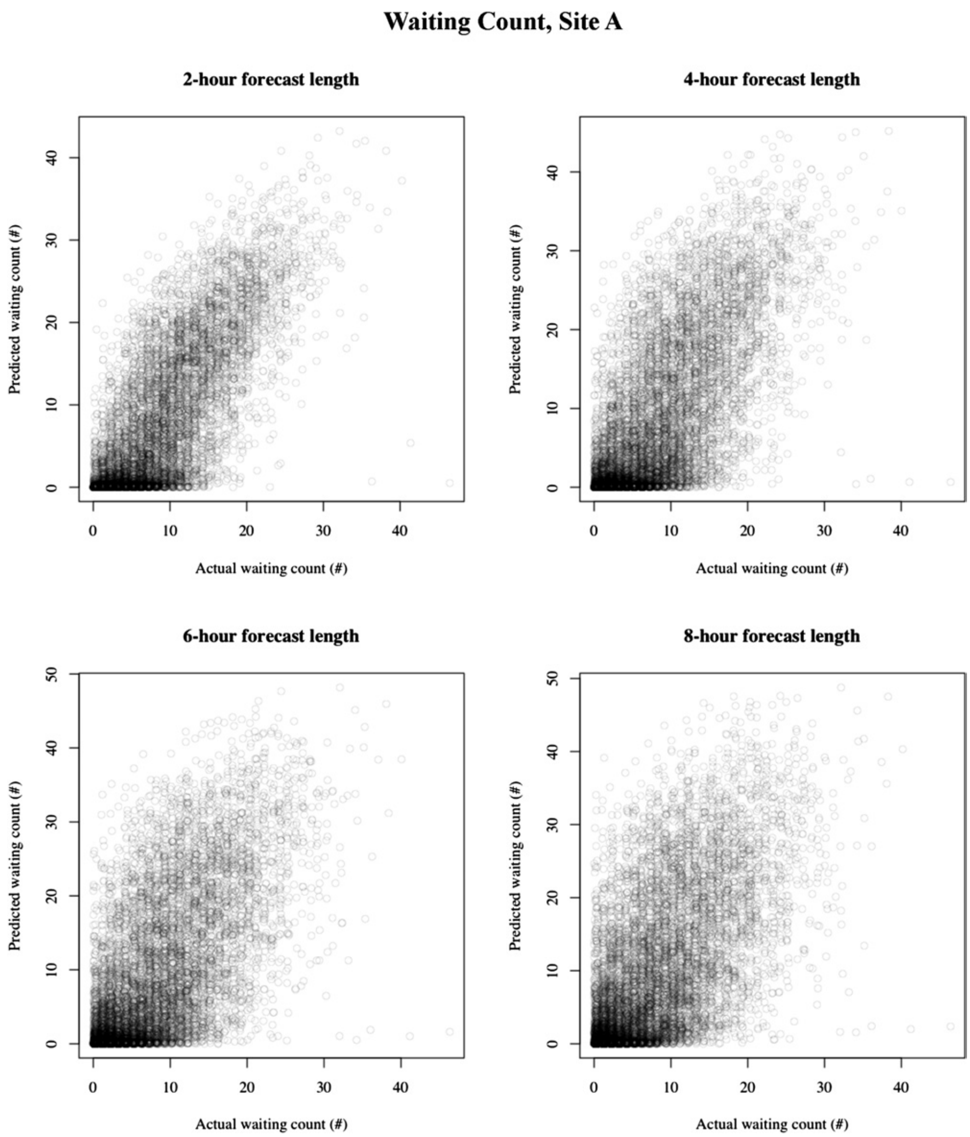

Figure E1. Actual (*x* axis) versus predicted (*y* axis) waiting counts at site A. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The waiting count is expressed as the total number of patients in the waiting room at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

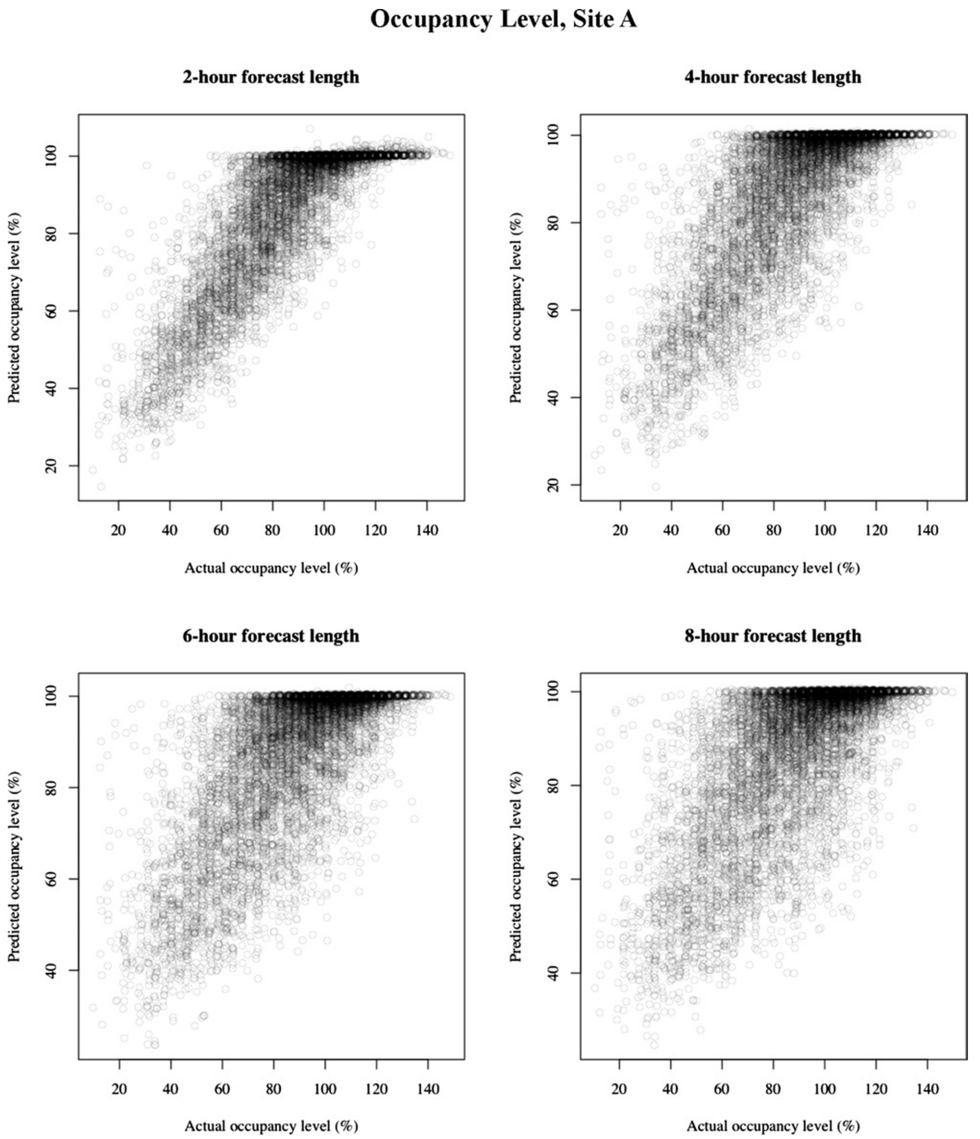

Figure E2. Actual (*x* axis) versus predicted (*y* axis) occupancy levels at site A. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The occupancy level is expressed as the percentage of licensed beds that are filled at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

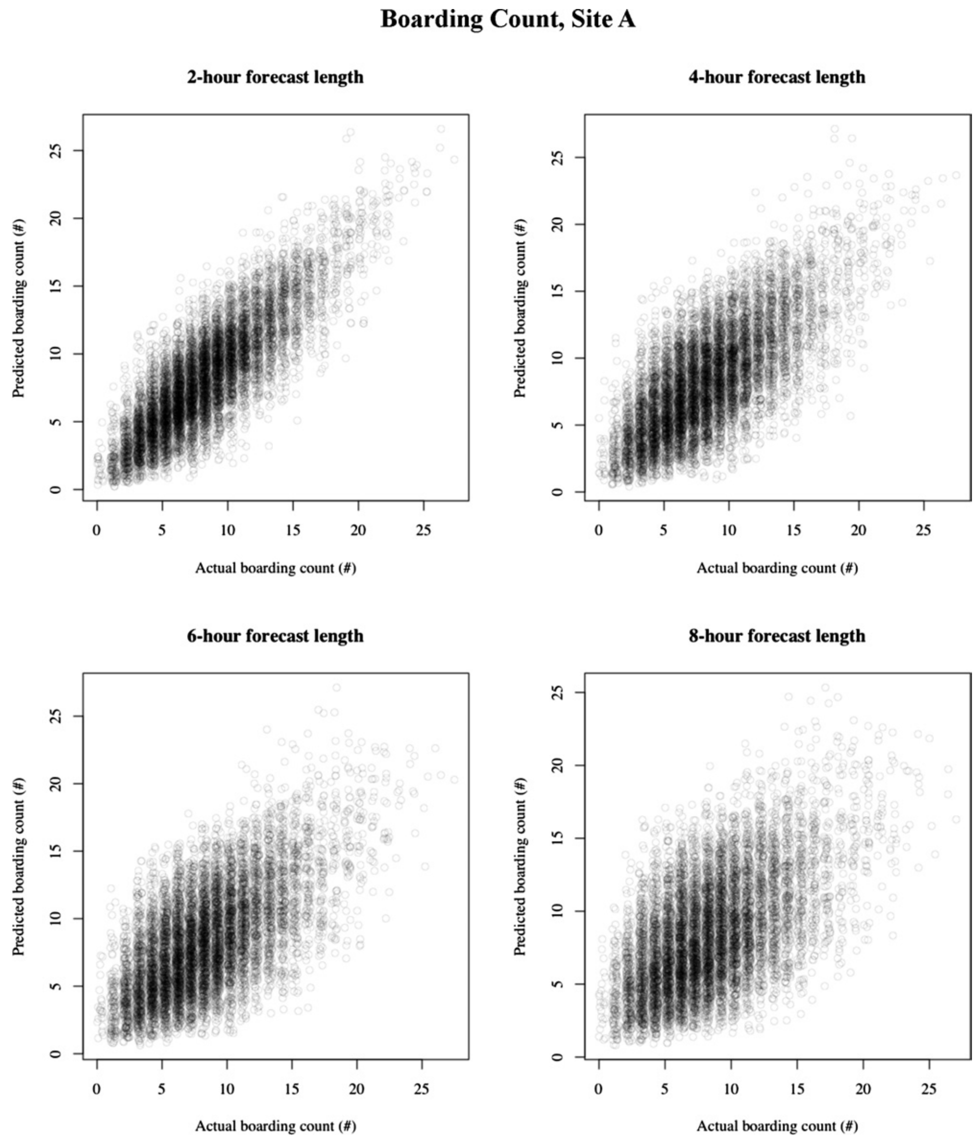

Figure E3. Actual (*x* axis) versus predicted (*y* axis) boarding counts at site A. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The boarding count is expressed as the number of patients with hospital admission orders who await inpatient beds at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

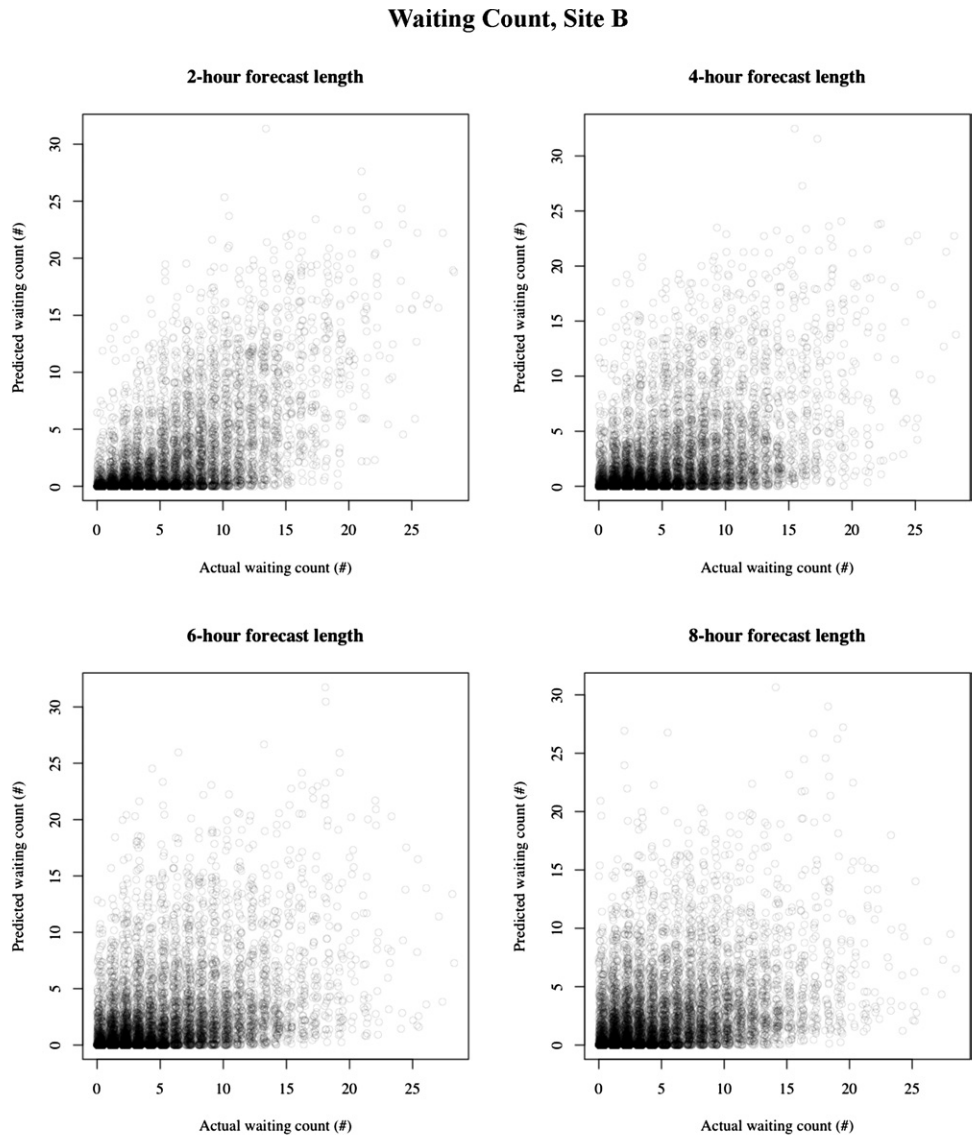

Figure E4. Actual (*x* axis) versus predicted (*y* axis) waiting counts at site B. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The waiting count is expressed as the total number of patients in the waiting room at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

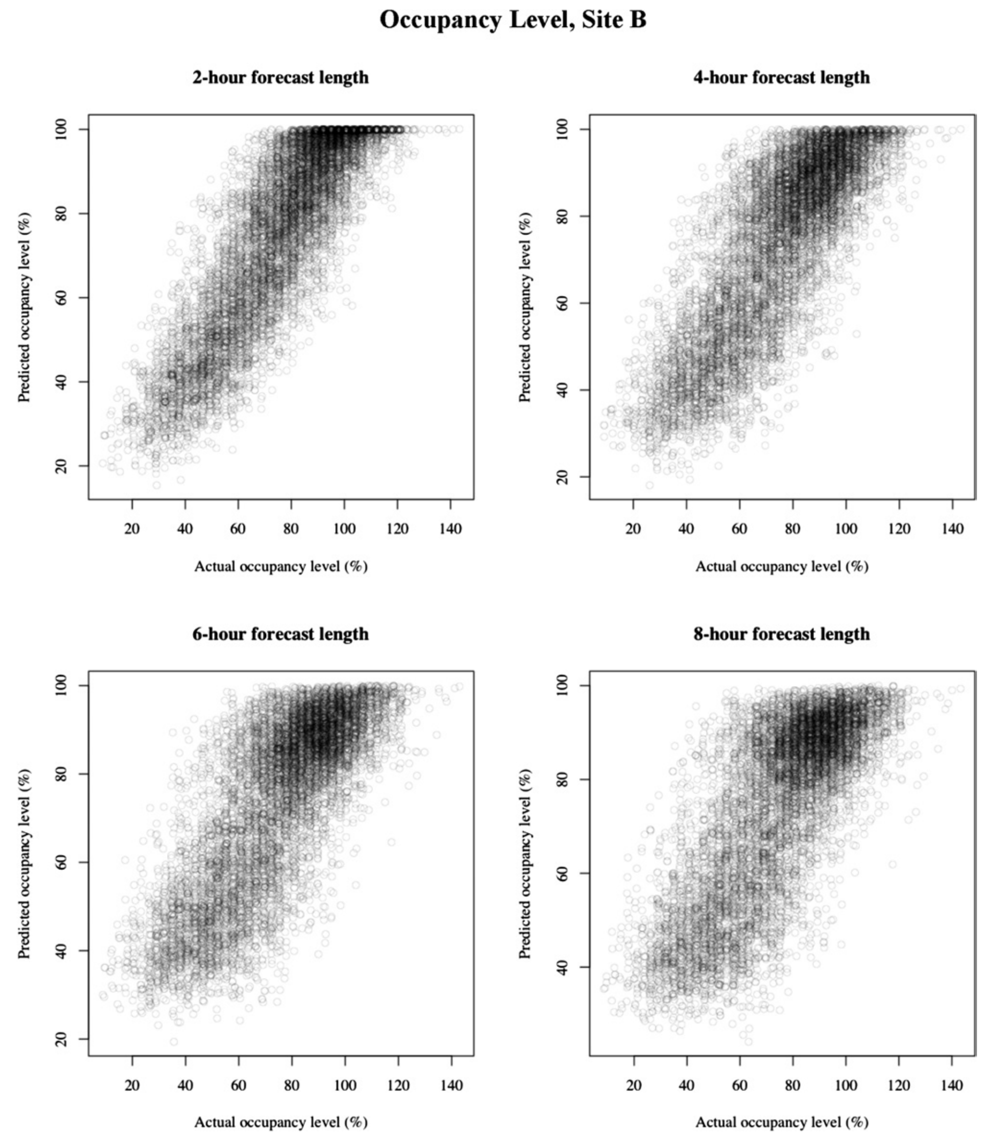

Figure E5. Actual (*x* axis) versus predicted (*y* axis) occupancy levels at site B. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The occupancy level is expressed as the percentage of licensed beds that are filled at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

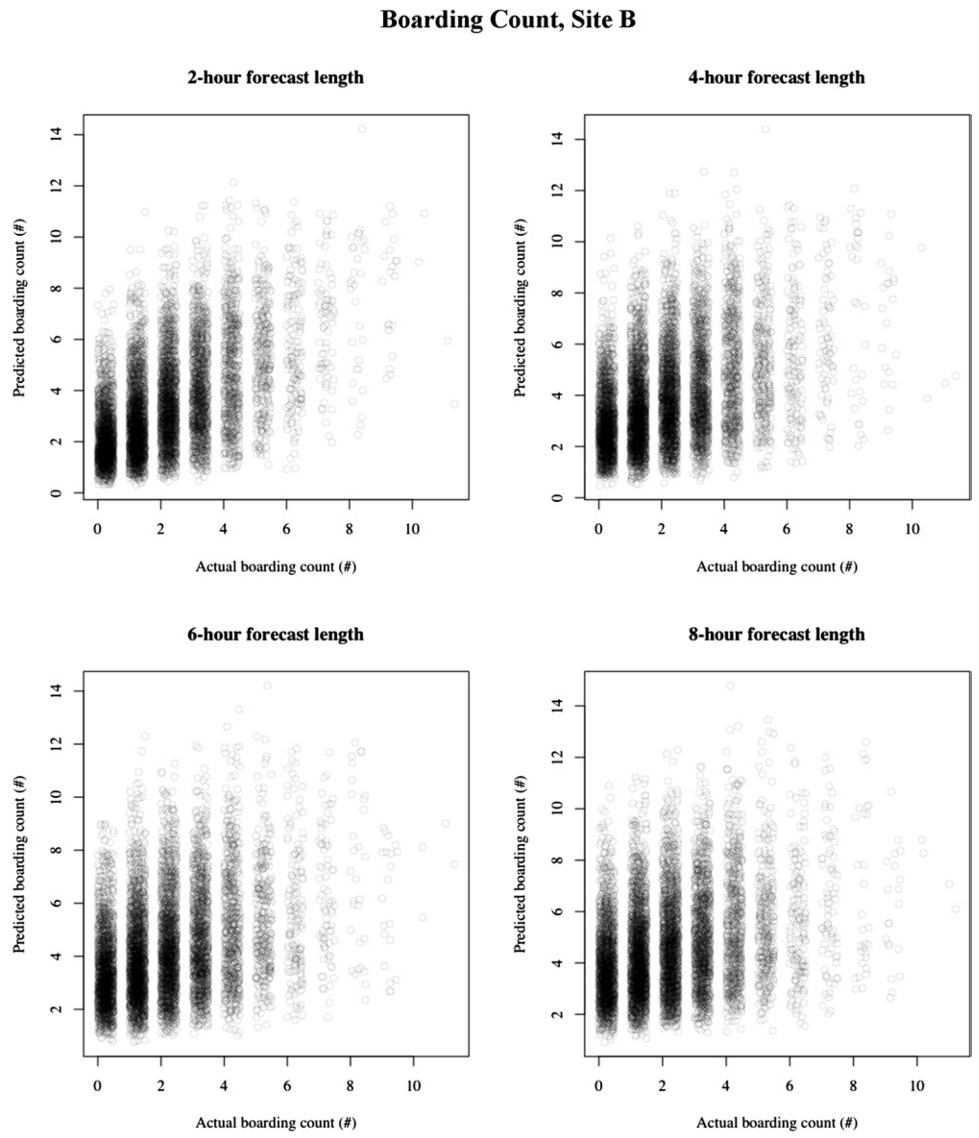

Figure E6. Actual (*x* axis) versus predicted (*y* axis) boarding counts at site B. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The boarding count is expressed as the number of patients with hospital admission orders who await inpatient beds at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

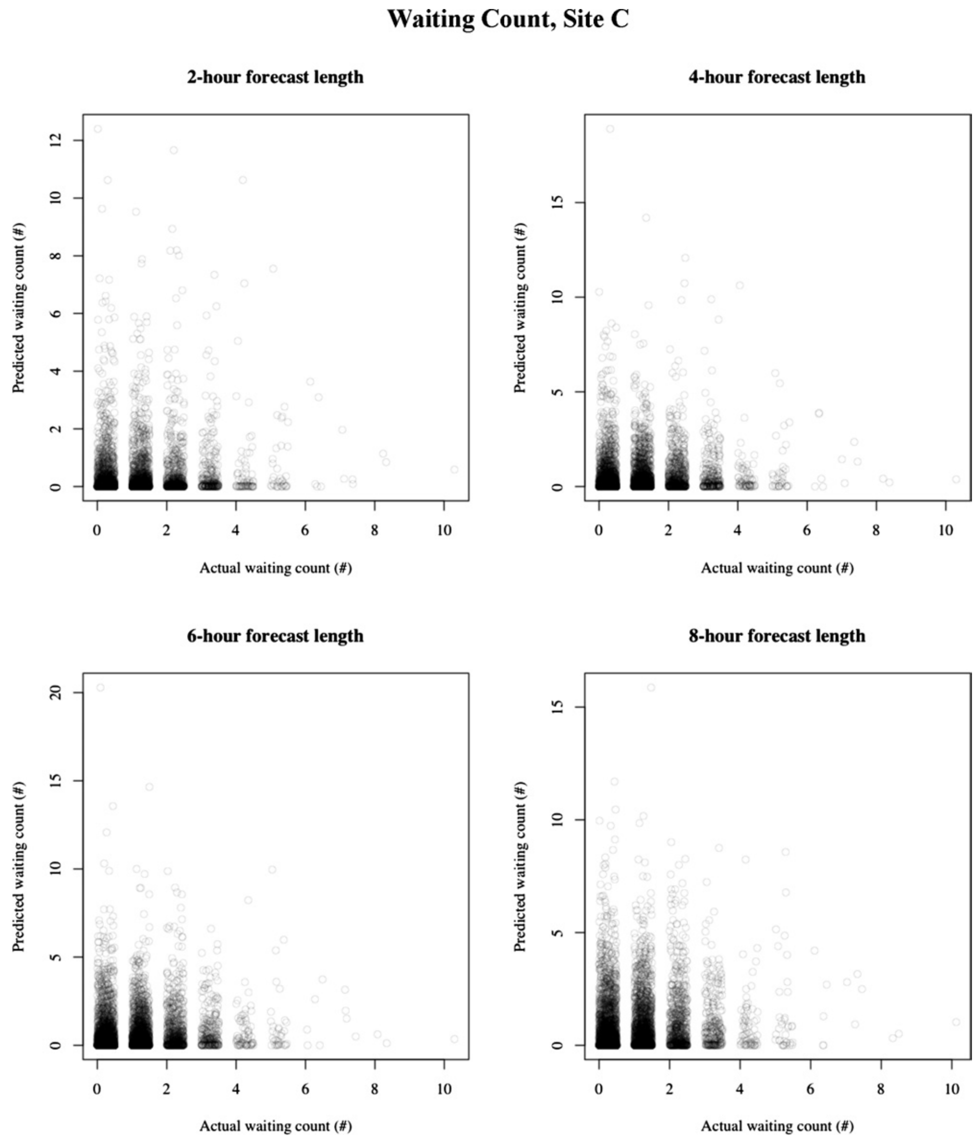

Figure E7. Actual (*x* axis) versus predicted (*y* axis) waiting counts at site C. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The waiting count is expressed as the total number of patients in the waiting room at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

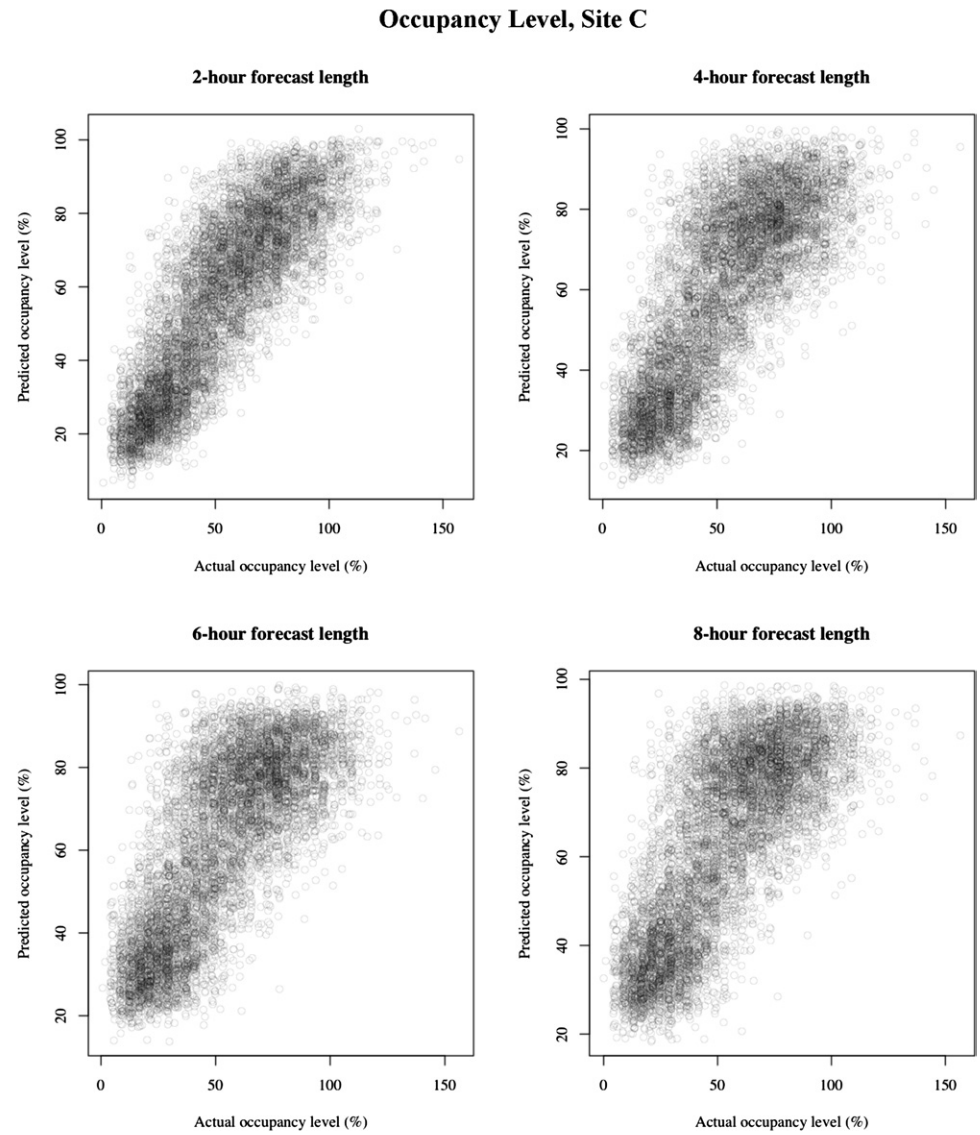

Figure E8. Actual (*x* axis) versus predicted (*y* axis) occupancy levels at site C. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The occupancy level is expressed as the percentage of licensed beds that are filled at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

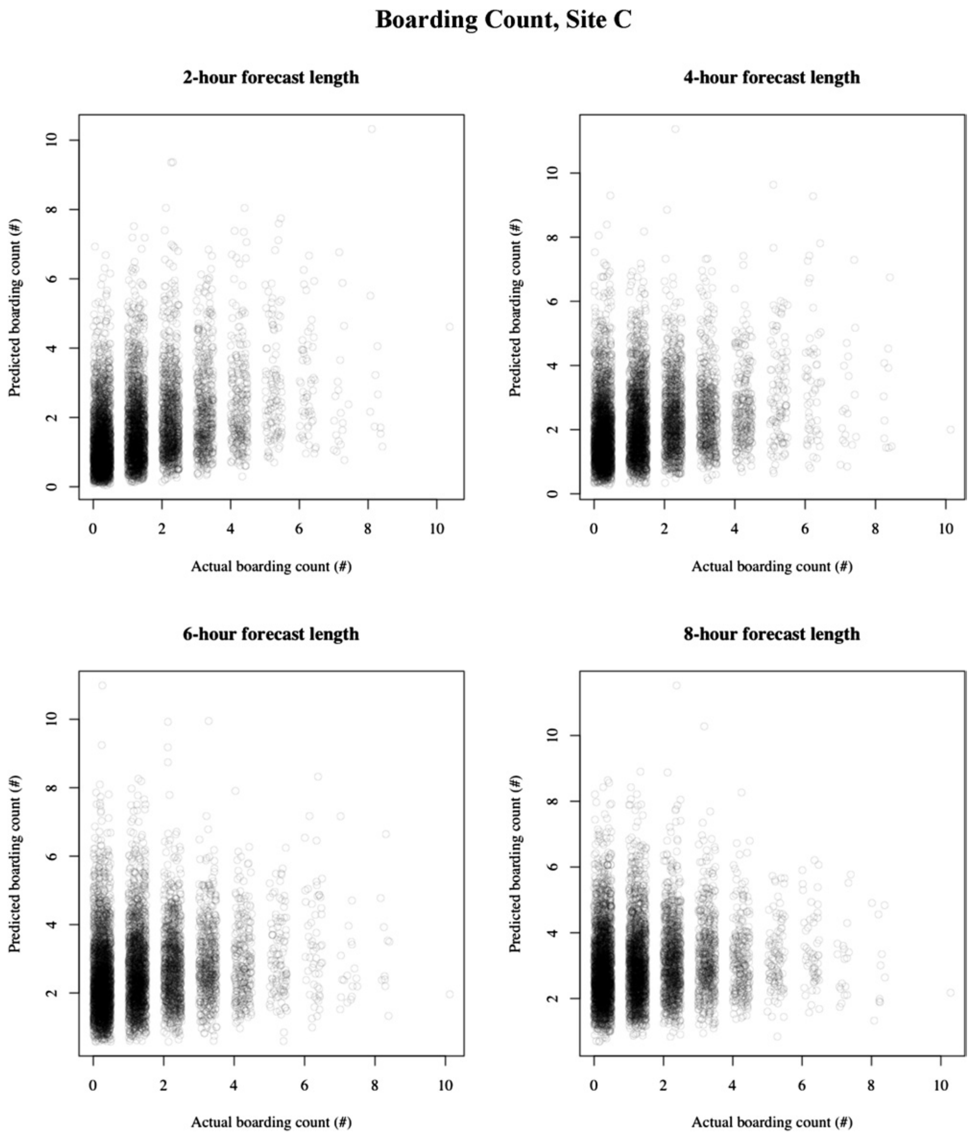

Figure E9. Actual (*x* axis) versus predicted (*y* axis) boarding counts at site C. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The boarding count is expressed as the number of patients with hospital admission orders who await inpatient beds at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

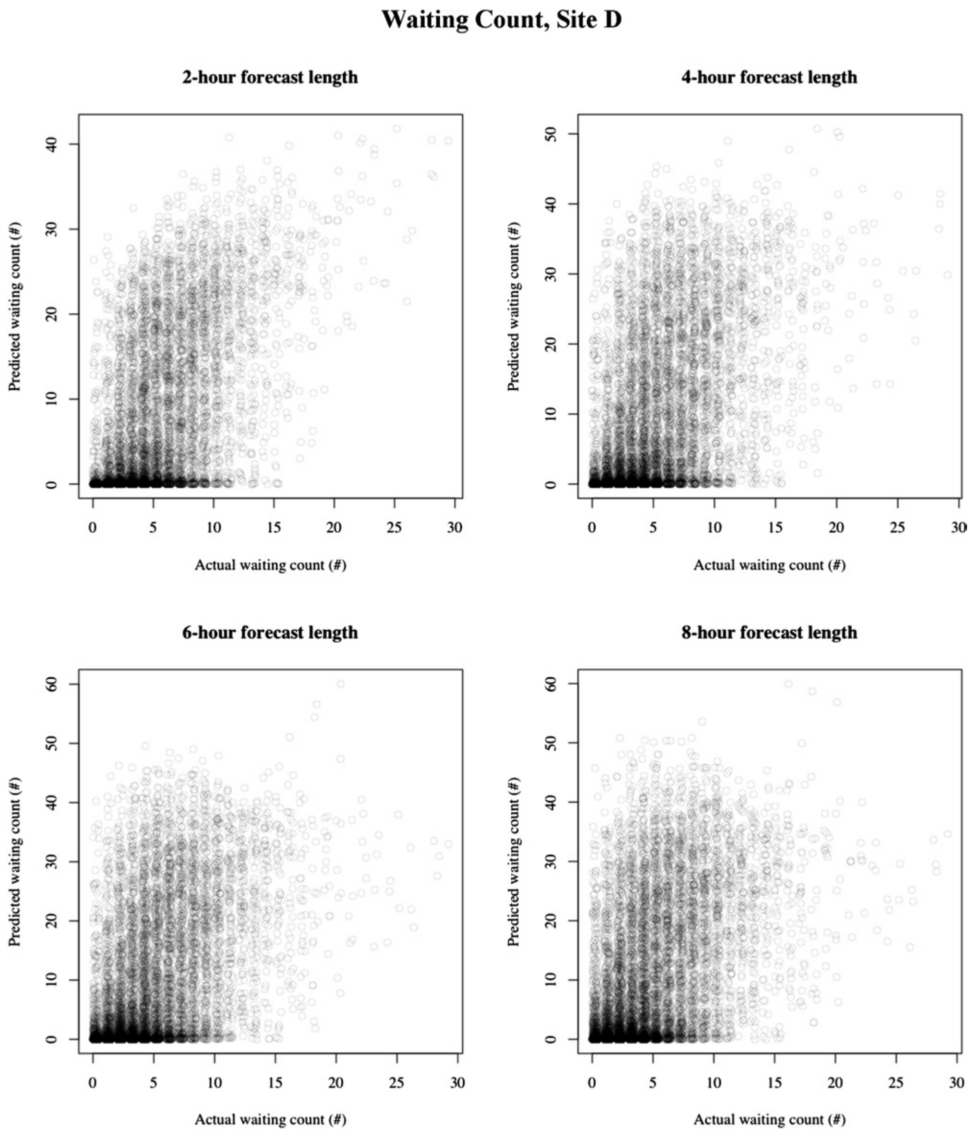

Figure E10. Actual (*x* axis) versus predicted (*y* axis) waiting counts at site D. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The waiting count is expressed as the total number of patients in the waiting room at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

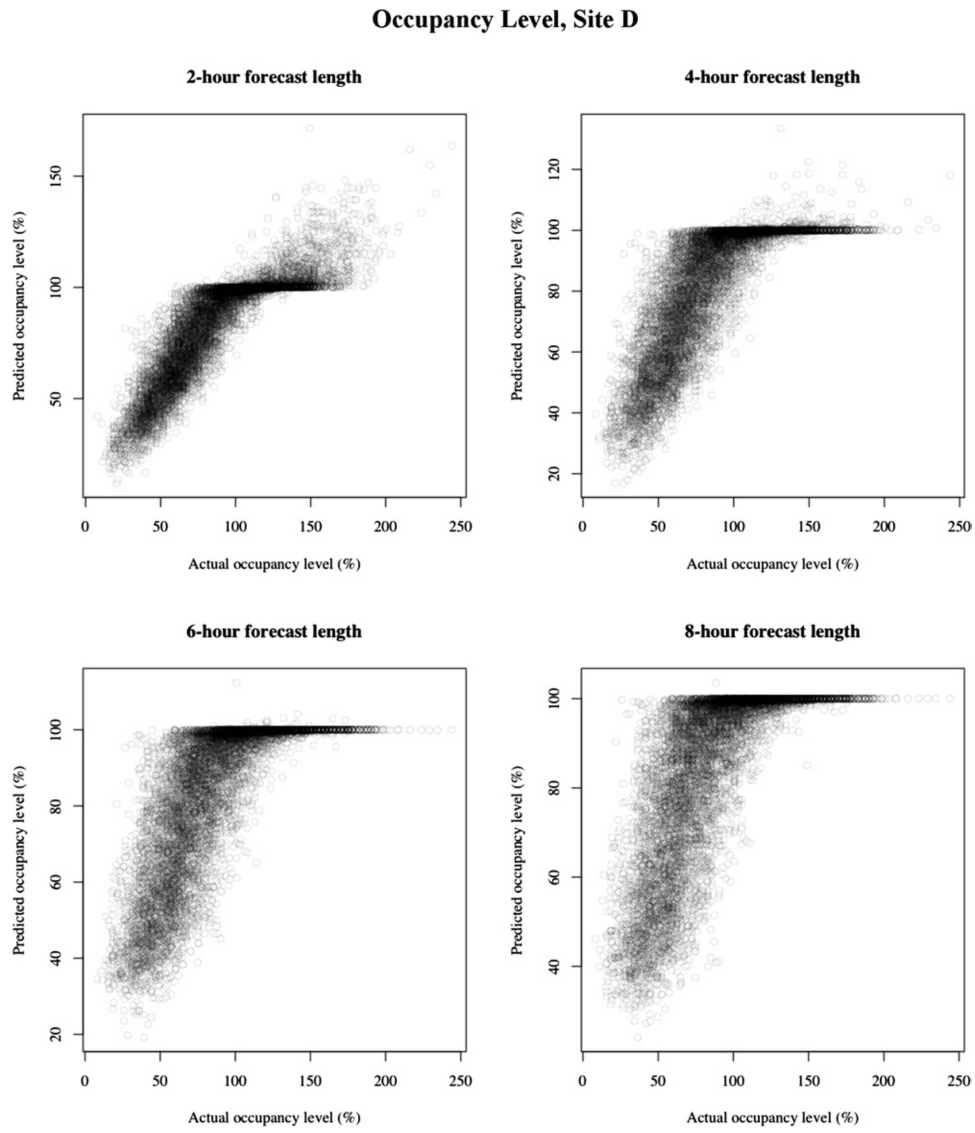

Figure E11. Actual (*x* axis) versus predicted (*y* axis) occupancy levels at site D. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The occupancy level is expressed as the percentage of licensed beds that are filled at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

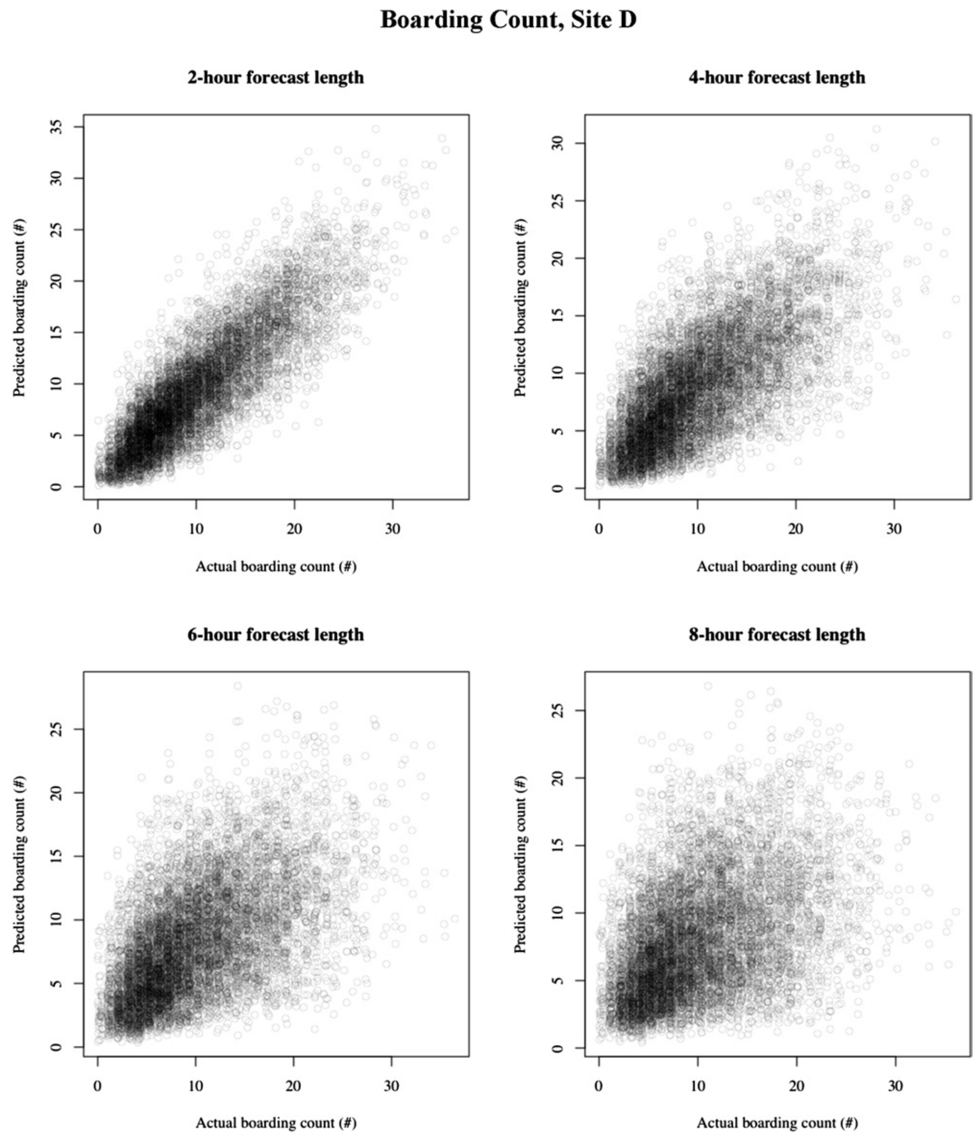

Figure E12. Actual (*x* axis) versus predicted (*y* axis) boarding counts at site D. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The boarding count is expressed as the number of patients with hospital admission orders who await inpatient beds at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

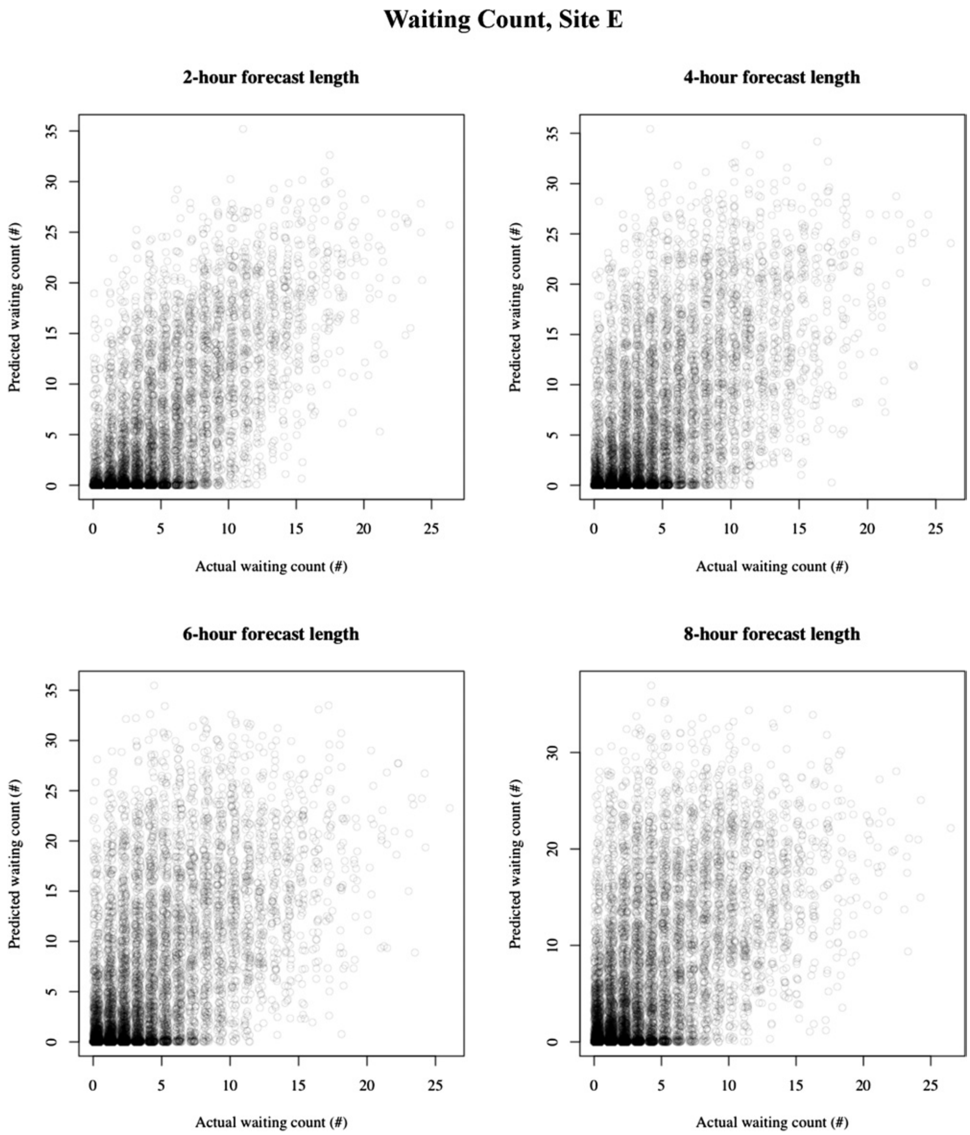

Figure E13. Actual (*x* axis) versus predicted (*y* axis) waiting counts at site E. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The waiting count is expressed as the total number of patients in the waiting room at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

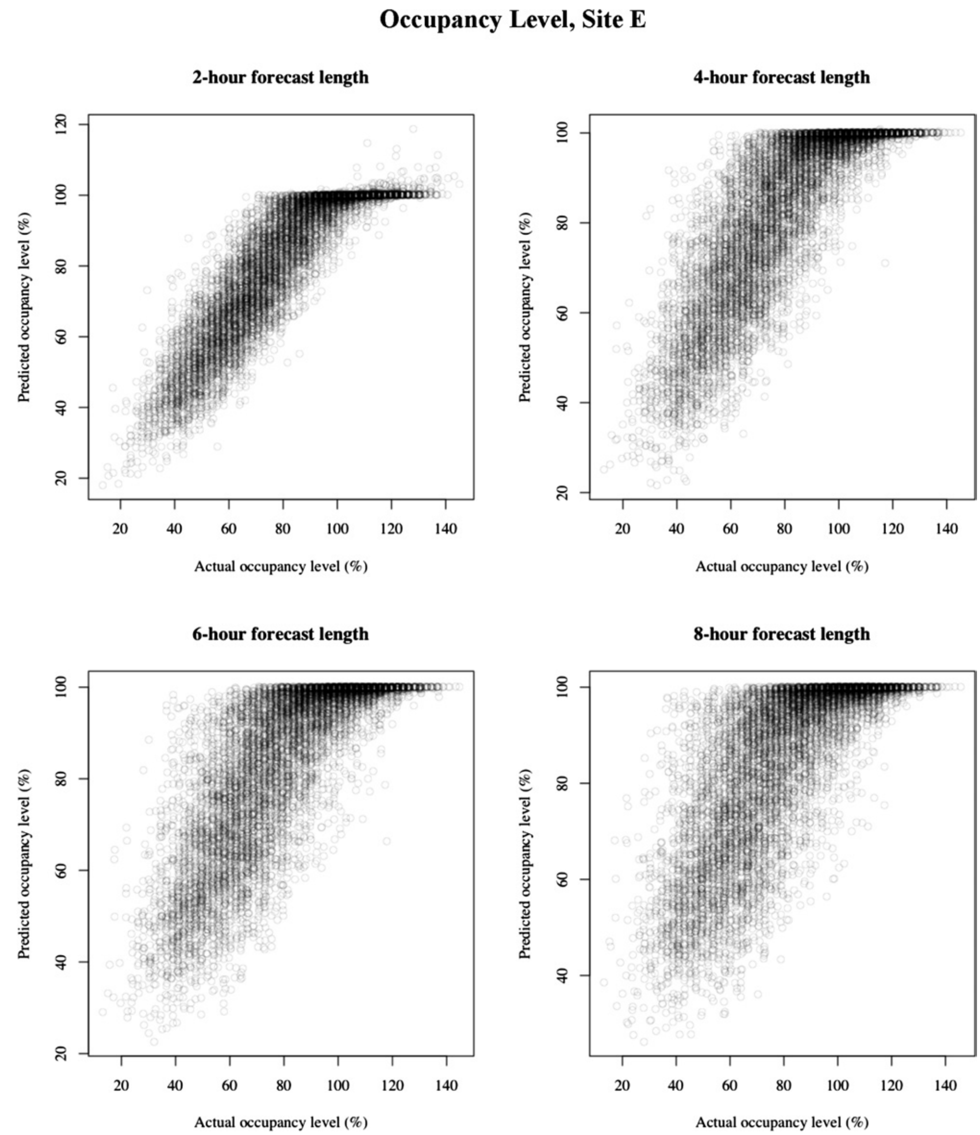

Figure E14. Actual (*x* axis) versus predicted (*y* axis) occupancy levels at site E. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The occupancy level is expressed as the percentage of licensed beds that are filled at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

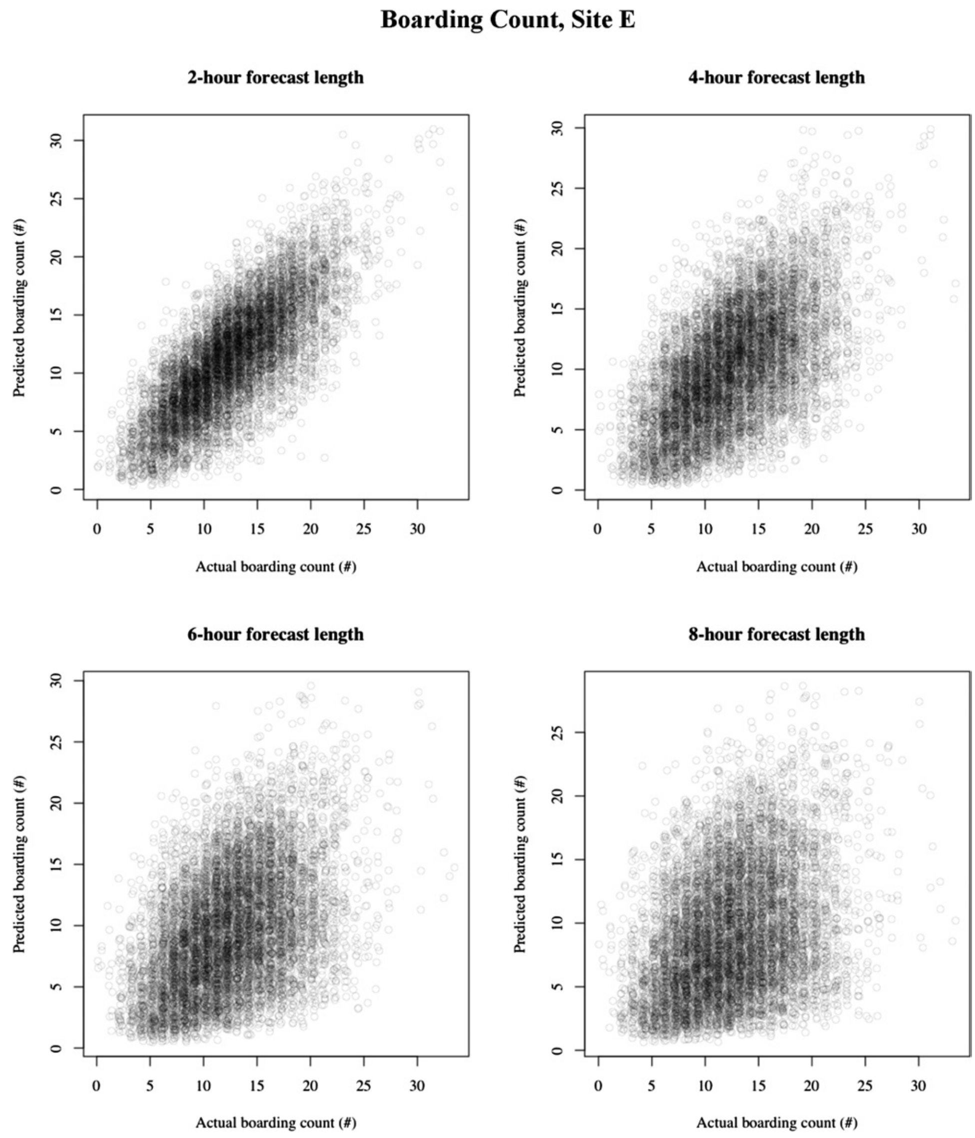

Figure E15. Actual (*x* axis) versus predicted (*y* axis) boarding counts at site E. The data represent forecast lengths of 2 hours (top left), 4 hours (top right), 6 hours (bottom left), and 8 hours (bottom right). The boarding count is expressed as the number of patients with hospital admission orders who await inpatient beds at one point. A small amount of uniformly distributed random noise was added to the *x* axis; this was necessary to improve legibility because observed patient counts are constrained to integers. The time-series resolution was reduced from 10-minute intervals to hourly intervals for the purpose of visualization.

APPENDIX E3.

Table E1. Reliability of the simulation versus autocorrelation in forecasting operational data at site A.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count				
Simulation (r)	0.79 (0.79 to 0.80)	0.71 (0.70 to 0.71)	0.65 (0.64 to 0.66)	0.61 (0.60 to 0.62)
Autocorrelation (r)	0.76 (0.76 to 0.76)	0.50 (0.49 to 0.51)	0.21 (0.20 to 0.22)	−0.05 (−0.06 to −0.04)
Occupancy level				
Simulation (r)	0.84 (0.83 to 0.84)	0.78 (0.77 to 0.78)	0.72 (0.72 to 0.73)	0.69 (0.68 to 0.69)
Autocorrelation (r)	0.78 (0.77 to 0.78)	0.52 (0.51 to 0.52)	0.27 (0.26 to 0.28)	0.10 (0.09 to 0.11)
Boarding count				
Simulation (r)	0.84 (0.84 to 0.84)	0.72 (0.72 to 0.73)	0.63 (0.63 to 0.64)	0.56 (0.55 to 0.57)
Autocorrelation (r)	0.85 (0.85 to 0.85)	0.73 (0.73 to 0.74)	0.65 (0.65 to 0.66)	0.59 (0.58 to 0.59)

*The Pearson coefficient of correlation is presented, with lower and upper bounds of the 95% CI in parentheses.

Table E2. Calibration of the simulation in forecasting operational data at site A.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	-0.4 ± 5.4	0.6 ± 6.9	1.4 ± 7.8	2.1 ± 8.3
Occupancy level (% of beds)	-0.1 ± 13.5	-0.3 ± 15.4	-0.4 ± 16.7	-0.5 ± 17.5
Boarding count (# of patients)	0.2 ± 2.3	0.2 ± 3.0	0.2 ± 3.4	0.3 ± 3.7

*The forecasting residuals are summarized, with the mean±standard deviation.

Table E3. Absolute error of the simulation in forecasting operational data at site A.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	2.7 (1.3-4.3)	3.1 (1.6-5.1)	3.4 (1.7-5.5)	3.5 (1.8-5.8)
Occupancy level (% of beds)	8.8 (4.1-15.0)	10.4 (4.9-17.5)	11.4 (5.4-19.2)	12.0 (5.7-20.3)
Boarding count (# of patients)	1.4 (0.7-2.5)	1.9 (0.9-3.2)	2.1 (1.0-3.6)	2.3 (1.1-3.9)

*The median absolute error is presented, with the interquartile range in parentheses.

Table E4. Reliability of the simulation versus autocorrelation in forecasting operational data at site B.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count				
Simulation (r)	0.66 (0.65 to 0.67)	0.52 (0.51 to 0.53)	0.43 (0.42 to 0.44)	0.39 (0.38 to 0.40)
Autocorrelation (r)	0.70 (0.69 to 0.70)	0.45 (0.44 to 0.46)	0.21 (0.20 to 0.21)	0.01 (0.00 to 0.02)
Occupancy level				
Simulation (r)	0.86 (0.85 to 0.86)	0.80 (0.80 to 0.81)	0.77 (0.77 to 0.78)	0.76 (0.75 to 0.76)
Autocorrelation (r)	0.77 (0.77 to 0.78)	0.46 (0.45 to 0.47)	0.14 (0.14 to 0.15)	−0.10 (−0.11 to −0.10)
Boarding count				
Simulation (r)	0.53 (0.53 to 0.54)	0.41 (0.40 to 0.41)	0.34 (0.33 to 0.35)	0.29 (0.28 to 0.30)
Autocorrelation (r)	0.63 (0.62 to 0.64)	0.48 (0.47 to 0.49)	0.40 (0.39 to 0.41)	0.36 (0.35 to 0.36)

*The Pearson coefficient of correlation is presented, with lower and upper bounds of the 95% CI in parentheses.

Table E5. Calibration of the simulation in forecasting operational data at site B.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	-2.6 ± 3.5	-2.5 ± 4.1	-2.4 ± 4.4	-2.3 ± 4.5
Occupancy level (% of beds)	0.1 ± 11.7	-0.4 ± 13.4	-0.1 ± 14.2	0.4 ± 14.7
Boarding count (# of patients)	1.5 ± 1.7	2.0 ± 1.9	2.4 ± 2.0	2.7 ± 2.1

*The forecasting residuals are summarized, with the mean±standard deviation.

Table E6. Absolute error of the simulation in forecasting operational data at site B.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	2.1 (1.1-3.2)	2.4 (1.3-3.7)	2.5 (1.4-3.9)	2.5 (1.4-4.0)
Occupancy level (% of beds)	7.7 (3.6-13.2)	9.0 (4.3-15.4)	9.6 (4.6-16.4)	10.0 (4.7-17.1)
Boarding count (# of patients)	0.9 (0.4-1.4)	1.0 (0.5-1.6)	1.0 (0.5-1.7)	1.0 (0.5-1.7)

*The median absolute error is presented, with the interquartile range in parentheses.

Table E7. Reliability of the simulation versus autocorrelation in forecasting operational data at site C.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count				
Simulation (r)	0.15 (0.13 to 0.16)	0.15 (0.14 to 0.16)	0.15 (0.14 to 0.16)	0.15 (0.14 to 0.17)
Autocorrelation (r)	0.15 (0.14 to 0.16)	0.08 (0.06 to 0.09)	−0.01 (−0.02 to 0.00)	−0.05 (−0.06 to −0.04)
Occupancy level				
Simulation (r)	0.83 (0.83 to 0.84)	0.77 (0.77 to 0.77)	0.75 (0.75 to 0.76)	0.75 (0.74 to 0.75)
Autocorrelation (r)	0.72 (0.71 to 0.72)	0.37 (0.36 to 0.37)	0.02 (0.01 to 0.03)	−0.27 (−0.28 to −0.27)
Boarding count				
Simulation (r)	0.36 (0.35 to 0.37)	0.26 (0.25 to 0.27)	0.17 (0.16 to 0.18)	0.11 (0.10 to 0.12)
Autocorrelation (r)	0.30 (0.29 to 0.30)	0.17 (0.16 to 0.18)	0.04 (0.03 to 0.04)	−0.07 (−0.08 to −0.06)

*The Pearson coefficient of correlation is presented, with lower and upper bounds of the 95% CI in parentheses.

Table E8. Calibration of the simulation in forecasting operational data at site C.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	−0.4 ± 1.1	−0.3 ± 1.2	−0.1 ± 1.3	0.0 ± 1.4
Occupancy level (% of beds)	4.6 ± 14.6	7.0 ± 16.9	8.9 ± 17.4	10.6 ± 17.5
Boarding count (# of patients)	0.5 ± 1.4	1.0 ± 1.5	1.5 ± 1.6	1.9 ± 1.6

*The forecasting residuals are summarized, with the mean±standard deviation.

Table E9. Absolute error of the simulation in forecasting operational data at site C.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	0.6 (0.4-0.7)	0.6 (0.4-0.7)	0.6 (0.4-0.7)	0.6 (0.4-0.7)
Occupancy level (% of beds)	9.6 (4.5-16.5)	11.1 (5.2-19.1)	11.5 (5.4-20.0)	11.6 (5.4-20.1)
Boarding count (# of patients)	0.8 (0.4-1.2)	0.9 (0.4-1.2)	0.9 (0.4-1.2)	1.0 (0.4-1.2)

*The median absolute error is presented with the interquartile range in parentheses.

Table E10. Reliability of the simulation versus autocorrelation in forecasting operational data at site D.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count				
Simulation (r)	0.63 (0.63-0.64)	0.56 (0.56-0.57)	0.53 (0.52-0.53)	0.50 (0.50-0.51)
Autocorrelation (r)	0.65 (0.64-0.66)	0.41 (0.40-0.42)	0.19 (0.18-0.20)	0.03 (0.02-0.03)
Occupancy level				
Simulation (r)	0.87 (0.87-0.87)	0.79 (0.79-0.79)	0.76 (0.75-0.76)	0.73 (0.73-0.73)
Autocorrelation (r)	0.86 (0.86-0.86)	0.61 (0.61-0.62)	0.31 (0.30-0.32)	0.03 (0.02-0.04)
Boarding count				
Simulation (r)	0.85 (0.85-0.86)	0.70 (0.70-0.71)	0.55 (0.54-0.55)	0.43 (0.42-0.43)
Autocorrelation (r)	0.84 (0.84-0.84)	0.65 (0.64-0.65)	0.45 (0.44-0.45)	0.27 (0.26-0.28)

*The Pearson coefficient of correlation is presented, with lower and upper bounds of the 95% CI in parentheses.

Table E11. Calibration of the simulation in forecasting operational data at site D.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	2.0 ± 7.2	3.6 ± 9.2	4.9 ± 10.2	6.0 ± 10.6
Occupancy level (% of beds)	−5.3 ± 19.8	−5.6 ± 23.5	−4.6 ± 24.8	−3.5 ± 25.8
Boarding count (# of patients)	−0.4 ± 3.2	−1.2 ± 4.4	−1.7 ± 5.2	−2.0 ± 5.7

*The forecasting residuals are summarized, with the mean±standard deviation.

Table E12. Absolute error of the simulation in forecasting operational data at site D.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	1.9 (0.9-3.0)	2.0 (1.0-3.2)	2.0 (1.0-3.3)	2.1 (1.0-3.4)
Occupancy level (% of beds)	11.6 (5.4-20.1)	14.5 (6.9-25.0)	15.8 (7.5-27.3)	16.8 (8.0-28.7)
Boarding count (# of patients)	2.0 (0.9-3.4)	2.8 (1.3-4.7)	3.4 (1.6-5.7)	3.8 (1.9-6.3)

*The median absolute error is presented, with the interquartile range in parentheses.

Table E13. Reliability of the simulation versus autocorrelation in forecasting operational data at site E.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count				
Simulation (r)	0.73 (0.73 to 0.74)	0.65 (0.64 to 0.66)	0.60 (0.59 to 0.61)	0.57 (0.57 to 0.58)
Autocorrelation (r)	0.70 (0.69 to 0.70)	0.44 (0.44 to 0.45)	0.18 (0.17 to 0.18)	−0.06 (−0.07 to −0.05)
Occupancy level				
Simulation (r)	0.89 (0.89 to 0.89)	0.83 (0.83 to 0.83)	0.78 (0.78 to 0.79)	0.75 (0.75 to 0.75)
Autocorrelation (r)	0.85 (0.85 to 0.85)	0.60 (0.60 to 0.61)	0.31 (0.30 to 0.32)	0.04 (0.03 to 0.05)
Boarding count				
Simulation (r)	0.76 (0.76 to 0.76)	0.57 (0.57 to 0.58)	0.43 (0.42 to 0.43)	0.33 (0.32 to 0.34)
Autocorrelation (r)	0.80 (0.80 to 0.81)	0.68 (0.68 to 0.68)	0.58 (0.57 to 0.58)	0.50 (0.49 to 0.51)

*The Pearson coefficient of correlation is presented, with lower and upper bounds of the 95% CI in parentheses.

Table E14. Calibration of the simulation in forecasting operational data at site E.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	0.4 ± 4.3	1.2 ± 5.2	2.0 ± 5.8	2.7 ± 6.2
Occupancy level (% of beds)	1.4 ± 10.9	2.6 ± 13.3	3.7 ± 14.7	4.6 ± 15.7
Boarding count (# of patients)	−1.1 ± 3.3	−2.0 ± 4.4	−2.6 ± 5.2	−3.1 ± 5.7

*The forecasting residuals are summarized, with the mean±standard deviation.

Table E15. Absolute error of the simulation in forecasting operational data at site E.*

	2 Hours Ahead	4 Hours Ahead	6 Hours Ahead	8 Hours Ahead
Waiting count (# of patients)	1.5 (0.7-2.4)	1.6 (0.8-2.7)	1.7 (0.9-3.0)	1.7 (0.9-3.1)
Occupancy level (% of beds)	7.1 (3.4-12.2)	9.0 (4.3-15.4)	10.1 (4.8-17.3)	11.0 (5.2-18.6)
Boarding count (# of patients)	2.0 (1.0-3.5)	2.7 (1.3-4.5)	3.0 (1.4-5.0)	3.2 (1.5-5.3)

*The median absolute error is presented, with the interquartile range in parentheses.